

**Le, Emily**

---

**From:** Le, Emily  
**Sent:** Tuesday, August 03, 2004 9:45 AM  
**To:** Delaval, Jan  
**Subject:** RE: sequence search: 09/939537

-----Original Message-----

**From:** Delaval, Jan  
**Sent:** Tuesday, August 03, 2004 5:58 AM  
**To:** Le, Emily  
**Subject:** RE: sequence search: 09/939537

I need to know the exact points of fusion for each request., example number 3: 1-394 of # 29 fusion to position 1 of # 32.

-----Original Message-----

**From:** Le, Emily  
**Sent:** Monday, August 02, 2004 9:10 AM  
**To:** Delaval, Jan  
**Subject:** sequence search: 09/939537

Hi Jan,

Please provide a search of:

**[Le, Emily]** ~~1-Residues 1-395 of SEQ ID NO: 29~~

~~2-Residues 1-200 of SEQ ID NO: 31~~

3. Residues 1-394 of SEQ ID NO: 29 fused with SEQ ID NO: 32 **[Le, Emily]** , at position 1 of SEQ ID NO: 32

4. Residues 1-200 of SEQ ID NO: 31 fused with SEQ ID NO: 32 **[Le, Emily]** at position 1 of SEQ ID NO: 32

5. Residues 1-395 of SEQ ID NO: 29 fused with SEQ ID NO: 35 **[Le, Emily]** at position 1 of SEQ ID NO: 35

6. Residues 1-200 of SEQ ID NO: 31f used with SEQ ID NO: 35 **[Le, Emily]** at position 1 of SEQ ID NO: 35

7. Residues 1-394 of SEQ ID NO: 29 fused with SEQ ID NO: 32 **[Le, Emily]** at position 1 of SEQ ID NO: 32, fused with SEQ ID NO: 35 **[Le, Emily]** at position 1 of SEQ ID NO: 35

8. Residues 1-200 of SEQ ID NO: 31 fused with SEQ ID NO: 32 **[Le, Emily]** at position 1 of SEQ ID NO: 32, fused with SEQ ID NO: 35 **[Le, Emily]** at position 1 of SEQ ID NO: 35

for the above case. Please provide a printout of the first 100 hits. Thanks, Jan.

Emily Le  
Mailbox, Remsen 3C18  
Office, Remsen, 3C35  
(571) 272-0903



THIS PAGE BLANK (USPTO)



: Entered [jdelaval 3-Aug-04 11:02]

SE01  
MNRGVPFRHLLVQLALLPATQGNKVVLGKGDVTELTCTASQKSIQFHMKNNOIKILGNQGSFLT  
KQPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOLLVFGJLANSDTHLLOQOSLT  
LTLESPPSSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLONQKKEFKIDIVLAFQKASSI  
VYKKEGEVFEFPPLAFTEKLTGSGELMWOAEARASSSKWTFDLKNKEVSVKVTOPKLOMGKCLPL  
HLLFOALPOVYAGSGLTLALFAKTKGKLEHVEVNLVVRATOLOKNLTCEVWGPTSPKMLSLKLENKEAK  
VSKREKPYWVLNPEAGMOCCLLSDSQGVLESNIKVLPTWSTPVIH

: Entered [jdelaval 3-Aug-04 11:05]

SE02  
MNRGVPFRHLLVQLALLPATQGNKVVLGKGDVTELTCTASQKSIQFHMKNNOIKILGNQGSFLT  
KQPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOLLVFGJLANSDTHLLOQOSLT  
LTLESPPSSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLONQKKEFKIDIVL



; Entered [jdelaval 3-Aug-04 11:02]

SFO3  
MRBGVPFRHLLLVLOLALPPAATOGNKVVLGKKGDVTELTCTASQKKSIOFHMKNNSNOIKILGNOSFLT  
KGPSKUNDRADSRSLMDQGNFPLITKNLKTIEDSDTYICEVEDQKEVQLVFGLTANSDBTHLQGSFLT  
LTLESPPGSSPSVOCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLQNOQKVEFKIDIVLAFOKASSI  
VYKKGEOVEFSFPLAFVTEKLTGSGELMWOAERASSSKSWITFDLKNKEVSVKRYTODPKLQMGKULPL  
HLTLFOALPOVYAGSGNLTLLEAKTGKLEHQBENLVVMRATOLQKNLTCVMGPTSPKJMLSLKENKEAK  
VSKREKPVVNLNPEAGMOCILSDSGOVLLESNIKVLPTWSTPVEPKSCDKHTCPCPAPELGGPSVF  
LEPPKDKOTLMTSRTEVUTCVVVDVSHEDPEVKFNMYVDGVEHNAKTKPREEOYNSUTRYVVSULTVLAQ  
DMLNGKEYCKKVSNNKALPAPIEKTISKAKGQPREPOVYTTLPSSRDELTKNQVSLTCLVKGFTYPSDIAVEB  
ESNGQPENNYKTPPVLDSDGSFLYSKLTVDKSRMQQGNVPSCSVMEHALHNHYTQKSLSLSPGLQDE  
TCAEAQDGLDGLMTTDP1

; Entered [jdelaval 3-Aug-04 11:05]

SEO4  
MRBGVPFRHLLLVLOLALPPAATOGNKVVLGKKGDVTELTCTASQKKSIOFHMKNNSNOIKILGNOSFLT  
KGPSKUNDRADSRSLMDQGNFPLITKNLKTIEDSDTYICEVEDQKEVQLVFGLTANSDBTHLQGSFLT  
LTLESPPGSSPSVOCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLQNOQKVEFKIDIVPCPAPEPKSC  
DKHTCPELGGPSVFLFPPKDKOTLMTSRTEVUTCVVVDVSHEDPEVKFNMYVDGVEHNAKTKPREEO  
YNSUTRYVVSULTVLAQDMLNGKEYCKKVSNNKALPAPIEKTISKAKGQPREPOVYTTLPSSRDELTKNQVSL  
TCLVKGFTYPSDIAVEWESNGQPENNYKTPPVLDSDGSFLYSKLTVDKSRMQQGNVPSCSVMEHALHNH  
YTQKSLSLSPGLQDETCAEAQDGLDGLMTTDP1



: Entered [jdelaval 3-Aug-04 11:02]  
SPO5  
MRGVPFRHLLVLOLALLPATQGNKVVLGKGDVTELTCTASQKSIQFHMKNNOIKILGNQGSFLT  
KGPSKLNDRADSRSLMDQGNPLIITKNLKI ESDTYICEVEDQKEVQLVFGLTANS DTHLLOQSFLT  
LTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKQKVEFKIDI VPLAFOKASSI  
VYKKEGEQVEFSFPLAFTEKLTGSGELMWOAEARASSSKSWITFDLKNKEVSKRVTO DPKLQMGKXPL  
HLLTPOALPOVAGSGNLTAL EAKTGKLEHOEVLVVMRATOLQKNTCEVMGPTSPKMLSLKLEKXAK  
VSKREKPVVVLNBPAGMMOCLISDSGOVLLSENIKVLPTWSTPVHBRASALPAPPTGSALPDPQTASALP  
DPPASALPAAALAVISFLIGLIGVACVILARTRI

3

: Entered [jdelaval 3-Aug-04 11:05]  
SPO6  
MRGVPFRHLLVLOLALLPATQGNKVVLGKGDVTELTCTASQKSIQFHMKNNOIKILGNQGSFLT  
KGPSKLNDRADSRSLMDQGNPLIITKNLKI ESDTYICEVEDQKEVQLVFGLTANS DTHLLOQSFLT  
LTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKQKVEFKIDI VPRASALPAP  
TGSALPDPQTASALPDPASALPAAALAVISFLIGLIGVACVILARTRI

29 \$



; Entered [jdelaval 3-Aug-04 11:02]

SE07  
MNRGVPFRHLLVLOLALLPAATOGNKVVLGKKGDVLELTCTASOKKSIOPHMKNNOIKIILGNOGSFLT  
KGPSKLNDRADSRRLMDGNPPLIKNLKIEDSPYICEVEDQKEVOLLVFGLTANSDTHLLOGOSLT  
LTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELODGGTWTCTVLQNKKEVEFKIDI VLAFOQASIT  
VYKKEGEQVESFPPLAFVEKLTGSGELMWQERASSSSKSWITFDLKNKEVSVKRVTODPKLOMGKCLPL  
HLTLPOALPOVAGSGNLTLALAKTGKHLQEVNLVVMRATOLQKNLTCEVMGSPSKMLSLKENKEAK  
VSKREKPVVNLNPEAGMOCLLSDSGOVLLESNIKVLPTWSTFVPCPAPEPKSCDKHTCPELGGPSVF  
LEPPKDOTLMI SRPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKRREOYNSITRYVVSVLTVLHQ  
DMLNGKEYCKXSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELITGNQUSITGLVKGFTYPSDIAVEN  
BNGQEPNNYKITPPVLDSDGSFFLYSKLTVDKSRMGOGNVFSCSVMHEALHNHTYQKSLSLSPGLQDDE  
TCAEADGDELGLWTTDP1

; Entered [jdelaval 3-Aug-04 11:05]

SE08  
MNRGVPFRHLLVLOLALLPAATOGNKVVLGKKGDVLELTCTASOKKSIOPHMKNNOIKIILGNOGSFLT  
KGPSKLNDRADSRRLMDGNPPLIKNLKIEDSPYICEVEDQKEVOLLVFGLTANSDTHLLOGOSLT  
LTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELODGGTWTCTVLQNKKEVEFKIDI VPCPAPEPKSC  
DKHTCPELGGPSVFLPPKPKDOTLMI SRPEVTCVVVDVSHEDPEVKFNMYVDGVEVHNAKTKPREEQ  
YNSTYRVVSVLTVLHQDMLNGKEYCKXSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELITGNQUSL  
TCLVKGFTYPSDIAVENBNGQEPNNYKITPPVLDSDGSFFLYSKLTVDKSRMGOGNVFSCSVMHEALHNH  
YTOKSLSLSPGLQDDETCAEADGDELGLWTTDP1  
LAVISFLGLGLGACVLAATRI1



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 13:00:24 ; Search time 14.8924 Seconds  
(without alignments)  
4185.504 Million cell updates/sec

Title: SEQ3  
Perfect score: 3414  
Sequence: 1 MNRGVPRHLLVLQLALLP.....DETCAADQGEJLDTTP 648

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 125 summaries

Database : PIR 78:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2024	59.3	458	1 RWHUT4	T-cell surface gly
2	1876	55.0	432	1 RWC2T4	T-cell surface gly
3	1738	50.9	432	1 RMCQT4	T-cell surface gly
4	1266	37.1	330	1 GHNU	Ig gamma-1 chain C
5	1254	36.7	255	4 S31866	Ig gamma-1 chain C
6	1252.5	36.7	374	2 S69339	Ig heavy chain V r
7	1218	35.7	234	2 PT0207	Ig gamma-1 chain C
8	1171	34.3	377	2 A23511	Ig gamma-3 chain C
9	1169	34.2	377	2 A60764	Ig gamma-3 chain C
10	1151	33.7	289	1 G3HUM1	Ig gamma-3 heavy C
11	1150	33.7	326	1 G2HU	Ig gamma-2 chain C
12	1149	33.7	459	2 A46254	CD4 precursor - ra
13	1141	33.4	327	1 G4HU	Ig gamma-4 chain C
14	1105	32.4	432	2 S30193	T-cell surface gly
15	999	29.3	457	2 A27449	T-cell surface gly
16	993	29.1	457	1 RMMST4	T-cell surface gly
17	947.5	27.8	398	1 G3MSM	Ig gamma-3 chain C
18	940.5	27.5	393	1 G1MSM	Ig gamma-1 chain C
19	922	27.0	323	1 GHRB	Ig gamma chain C r
20	915.5	26.8	399	1 G2MSAM	Ig gamma-2a chain
21	911.5	26.7	328	2 I47160	Ig gamma-2b chain
22	911.5	26.7	328	2 I47159	Ig gamma-2a chain
23	906.5	26.6	277	2 I47162	Ig gamma-4 chain C
24	906.5	26.6	329	1 G2GP	Ig gamma-2 chain C
25	891.5	26.1	328	2 I47158	Ig gamma-1 chain C
26	884.5	25.9	405	1 G2MSBM	Ig gamma-2b chain
27	884.5	25.9	470	2 S22080	Ig heavy chain pre
28	883.5	25.9	328	2 I47161	Ig gamma-3 chain C
29	861	25.2	472	2 S31459	Ig gamma-1 chain -

30	851	24.9	308	2 C30554	Ig heavy chain C r
31	849.5	24.9	329	1 G3MSC	Ig gamma-3 chain C
32	844	24.7	333	2 PS0018	Ig gamma-2b chain
33	839.5	24.6	444	2 PC4436	monoclonal antibody
34	836.5	24.5	469	2 S37483	Ig gamma-2a chain
35	828	24.3	326	2 PS0017	Ig gamma-1 chain C
36	826.5	24.2	324	1 G1MS	Ig gamma-1 chain C
37	817.5	23.9	329	2 S00847	Ig gamma-2c chain
38	816	23.9	335	1 G2MSAB	Ig gamma-2a chain
39	814.5	23.9	330	1 G2MSA	Ig gamma-2a chain
40	812	23.8	446	2 S40295	Ig gamma-2a chain
41	811	23.8	474	1 G2MS11	Ig gamma-2a chain
42	795	23.3	322	2 PS0019	Ig gamma-2a chain
43	782.5	22.9	475	2 S01321	Ig gamma-2b chain
44	774	22.7	327	2 S06611	Ig gamma-2 chain C
45	702	20.6	180	2 I46732	Ig gamma heavy cha
46	574.5	16.8	218	2 A36040	Ig heavy chain V-I
47	572.5	16.8	249	2 S69340	Ig heavy chain VH1
48	566	16.6	152	2 S14236	Ig gamma-1 chain C
49	394.5	11.6	572	2 B46529	Ig Y heavy chain (
50	390.5	11.4	627	2 S14683	Ig mu chain precu
51	389	11.4	549	2 S04845	Ig heavy chain pre
52	375.5	11.0	476	1 MHMSM	Ig mu chain C regi
53	371.5	10.9	592	2 S25705	Ig mu chain - shee
54	370	10.8	548	2 S38864	Ig epsilon chain C
55	364	10.7	455	1 MHMS	Ig mu chain C regi
56	364	10.7	455	1 A24976	Ig mu chain C regi
57	363.5	10.6	388	1 EHMS	Ig epsilon chain C
58	360	10.5	474	2 S15590	Ig heavy chain - h
59	359.5	10.5	343	2 S25644	Ig mu chain C regi
60	359.5	10.5	448	2 S03186	Ig heavy chain C r
61	359.5	10.5	454	1 MHY	Ig mu chain C regi
62	359	10.5	428	1 EHTU	Ig epsilon chain C
63	358.5	10.5	453	2 S37768	Ig mu chain C regi
64	358.5	10.5	479	1 MHRBM	Ig mu chain C regi
65	358	10.5	504	2 S00390	Ig gamma chain (c1
66	357	10.5	71	2 I60082	CD4 receptor - hum
67	355.5	10.4	429	1 EHRT	Ig epsilon chain C
68	353.5	10.4	391	1 MHRBT	Ig mu heavy chain
69	353	10.3	426	2 I36948	Ig epsilon-chain -
70	352.5	10.3	473	1 MHRUM	Ig mu chain C regi
71	349	10.2	458	1 MHRB	Ig mu chain C regi
72	348	10.2	452	1 MHNU	Ig mu chain C regi
73	348	10.2	457	2 S03961	Ig epsilon chain C
74	345	10.1	423	1 EHMS	Ig epsilon chain C
75	340.5	10.0	577	2 I50731	Ig heavy chain - n
76	340	10.0	450	1 MHG	Ig mu chain C regi
77	333.5	9.8	453	2 C31933	Ig gamma-3 chain C
78	331.5	9.7	112	2 B30503	Ig gamma-2a chain
79	330.5	9.7	684	2 S60266	Ig gamma-2a chain
80	313.5	9.2	433	2 S31436	Ig gamma-2a chain
81	305.5	8.9	99	2 S21461	Ig alpha chain C r
82	305	8.9	259	1 AHRB	Ig alpha chain C r
83	304.5	8.9	357	2 S09269	Ig alpha chain C r
84	302	8.8	338	2 S09276	Ig alpha chain C r
85	297	8.7	438	1 HVRK2	Ig mu chain C regi
86	296	8.7	342	2 I47175	Ig alpha chain C r
87	295.5	8.7	348	2 S09270	Ig alpha chain C r
88	295.5	8.7	585	2 A46507	Ig alpha chain - c
89	295	8.6	461	1 HVRKC0	Ig mu chain C regi
90	293.5	8.6	357	2 S09265	Ig alpha chain C r
91	293.5	8.6	438	1 HVRKCS	Ig mu chain C regi
92	292.5	8.6	347	2 S09274	Ig alpha chain C r
93	291	8.5	367	1 MHCH	Ig mu chain C regi
94	285.5	8.4	343	2 S09272	Ig alpha chain C r
95	284	8.3	244	2 S12328	Ig heavy chain C r
96	282.5	8.3	340	2 B22360	Ig alpha-2 chain C
97	281.5	8.2	393	1 HVRKCI	Ig mu chain C regi
98	280.5	8.2	99	2 S21462	T-cell surface gly
99	280.5	8.2	352	2 S09266	Ig alpha chain C r
100	280.5	8.2	353	1 ALHU	Ig alpha-1 chain C
101	279	8.2	339	2 S09264	Ig alpha chain C r
102	279	8.2	568	2 A45804	Ig mu chain C regi



```
103 278.5 8.2 474 2 150830 Ig mu chain - Lepi
104 278 8.1 352 2 505500 Ig alpha-1 chain C r
105 277.5 8.1 357 2 S09267 Ig alpha chain C r
106 276.5 8.1 340 2 156230 Ig alpha-2 chain -
107 275.5 8.1 88 2 A30503 Ig gamma-2b chain
108 275 8.1 568 2 A34891 Ig heavy chain pre
109 270.5 7.9 358 2 S09268 Ig alpha chain C r
110 269 7.9 360 2 S09271 Ig alpha chain C r
111 268 7.9 573 2 S12838 Ig mu chain precur
112 267 7.8 370 1 HVKCS Ig mu chain C regi
113 266.5 7.7 348 2 S09273 Ig alpha chain C r
114 263 7.7 340 1 A2HU Ig alpha-2 chain C
115 259.5 7.6 342 2 A45966 Ig alpha chain C r
116 257 7.5 111 2 S43148 Ig upsilon chain -
117 257 7.5 454 2 A46532 Ig mu chain C regi
118 249 7.3 335 2 S09275 Ig alpha chain C r
119 248.5 7.3 344 1 AAMS Ig alpha chain C r
120 245 7.2 1005 2 T18537 Ig heavy chain - c
121 243.5 7.1 220 2 C22360 Ig alpha-2 chain C
122 240.5 7.0 580 2 A46538 Ig heavy chain, se
123 211 6.2 448 2 A46533 Ig heavy chain C r
124 209 6.1 115 2 168731 IGE chain C4 regio
125 206 6.0 402 2 S20002 Ig heavy chain, se
```

## ALIGNMENTS

## RESULT 1

RHW04 T-cell surface glycoprotein CD4 precursor [validated] - human

A/Alternate names: T-cell surface antigen T4/Lew 3

C/Species: Homo sapiens (man)

C/Date: 28-May-1996 #sequence revision 31-Dec-1998 #text change 20-Apr-2001

C/Accession: A30872; A32722; A34194; A53287; I54176; I54257; A02109; A30039

R/Maddon, P.J.; Littman, D.R.; Godfrey, M.; Maddon, D.E.; Chess, L.; Axel, R.

Cell 42, 93-104, 1985

A/Title: The isolation and nucleotide sequence of a cDNA encoding the T cell surface pro

A/Reference number: A30872; MUID:85254948; PMID:2990730

A/Accession: A30872

A/Molecule type: mRNA

A/Residues: 1-25, 'N', 27-458 <MAD>

A/Experimental source: clone pT48

R/Littman, D.R.; Maddon, P.J.; Axel, R.

Cell 55, 541, 1988

A/Title: Corrected CD4 sequence.

A/Reference number: A30907; MUID:89028665; PMID:3263213

A/Contents: annotation; revision to residue 26

R/Camerini, D.; Seed, B.

Cell 60, 747-754, 1990

A/Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the vi

A/Accession: A32722; MUID:90182664; PMID:2107024

A/Accession: A32722

A/Status: nucleic acid sequence not shown; not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 26-426, 428-458 <CAM>

R/Carr, S.A.; Hemling, M.E.; Folsena-Messerman, G.; Sweet, R.W.; Annumula, K.; Barr, J.R.;

J. Biol. Chem. 264, 21286-21295, 1989

A/Title: Protein and carbohydrate structural analysis of a recombinant soluble CD4 recep

A/Reference number: A34194; MUID:90078232; PMID:2592374

A/Contents: disulfide bonds; carbohydrate-binding sites

A/Accession: A34194

A/Molecule type: protein

A/Residues: 26-394 <CAR>

R/Lederman, S.; Demattino, J.A.; Daugherty, B.L.; Foeldvari, I.; Yellin, M.J.; Cleary, A

Mol. Immunol. 28, 1171-1181, 1991

A/Title: A single amino acid substitution in a common African allele of the CD4 molecule

A/Reference number: A53287; MUID:92072595; PMID:1961196

A/Accession: A53287

A/Status: not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 250-264, 'W', 266-280 <LED>

A/Note: sequence extracted from NCBI backbone (NCBIP:68249)

R/Edwards, M.C.; Gibbs, R.A.

Genomics 14, 590-597, 1992

A/Title: A human dimorphism resulting from loss of an Alu.

A/Reference number: I54176; MUID:9305387; PMID:1330888

A/Accession: I54176

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-72 <RES>

R/Hodge, T.W.; Sasso, D.R.; McDougal, J.S.

Hum. Immunol. 30, 99-104, 1991

A/Title: Humans with OKT4-epitope deficiency have a single nucleotide base change in the

A/Reference number: I54297; MUID:91216786; PMID:1708753

A/Accession: I54297

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-264, 'W', 266-458 <RES2>

A/Cross-references: GB:M35160; NID:g179143; PIDN:AAA16069.1; PID:g179144

C/Comment: Macrophage tropic strains of HIV-1 bind to a complex of chemokine (C-C) recep

C/Genetics:

A/Gene: GDB:CD4

A/Map position: 12pter-12p12

A/Introns: 16/3

C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology

C/Keywords: AIDS; duplication; glycoprotein; T-cell; transmembrane protein

F/1-25/Domain: signal sequence #status predicted <SIG>

F/26-458/Product: T-cell surface glycoprotein CD4 #status experimental <MAT>

F/34-111/Domain: immunoglobulin homology <IM1>

F/136-186/Domain: immunoglobulin homology #status atypical <IM2>

F/216-299/Domain: immunoglobulin homology <IM3>

F/321-372/Domain: immunoglobulin homology <IM4>

F/397-420/Domain: transmembrane #status predicted <TM>

F/431-458/Domain: intracellular #status predicted <INT>

F/41-109,155-184,328-370/Disulfide bonds: #status experimental

F/236,325/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 59.3%; Score 2024; DB 1; Length 458;

Best Local Similarity 99.2%; Pred. No. 1,7e-112;

Matches 393; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 MNRGVPFRHLLVQLALPAATQGNKRVIGKKDVELTCTASQKKSIOFHMNSNOIK 60
DB 1 MNRGVPFRHLLVQLALPAATQGNKRVIGKKDVELTCTASQKKSIOFHMNSNOIK 60
QY 61 ILNQGSEFLTKGSPSKNDPRADSRSLMDQGNFPLIKNLTEDSDTYICEVEDQKEEVOL 120
DB 61 ILNQGSEFLTKGSPSKNDPRADSRSLMDQGNFPLIKNLTEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 LVFGLTANSDTHLLQGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
QY 181 TWTCVTVLQNGKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
DB 181 TWTCVTVLQNGKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
QY 241 QARRASSKSWITFDLKNKEVSIVKRYTQDPKLGKGLPHLTLPLQALPOYASSGNITLA 300
DB 241 QARRASSKSWITFDLKNKEVSIVKRYTQDPKLGKGLPHLTLPLQALPOYASSGNITLA 300
QY 301 LEAKTGKLEHVEVLVVMRATQLOKNLTCEWVGTPSLMLSTLKENEAIVSKREKPVWV 360
DB 301 LEAKTGKLEHVEVLVVMRATQLOKNLTCEWVGTPSLMLSTLKENEAIVSKREKPVWV 360
QY 361 LNEBAGMOCLLSDSGQVLLSNIKVLPWTSTVQV 396
DB 361 LNEBAGMOCLLSDSGQVLLSNIKVLPWTSTVQV 396
```

## RESULT 2

RWC274

T-cell surface glycoprotein CD4 - chimpanzee



N/Alternate names: T-cell surface antigen T4/Len 3  
C/Species: Pan troglodytes (chimpanzee)  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999  
R/Accession: B32722; A46534  
R/Camerini, D.; Seed, B.  
Cell 60, 747-754, 1990  
A/Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the vi  
A/Reference number: A32722; MUID:90182664; PMID:2107024  
A/Accession: B32722  
A/Molecule type: mRNA  
A/Residues: 1-432 <CAM>  
A/Cross-references: GB:M31135  
R/Fomsgaard, A.; Hirsch, V.M.; Johnson, P.R.  
Eur. J. Immunol. 22, 2973-2981, 1992  
A/Title: Cloning and sequences of primate CD4 molecules: diversity of the cellular recep  
A/Reference number: A46534; MUID:93049640; PMID:1425921  
A/Accession: A46534  
A/Molecule type: not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 3-399 <POM>  
A/Note: sequence extracted from NCBI backbone (NCBI:118332)  
C/Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells  
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
C/Keywords: duplication; glycoprotein; T-cell; transmembrane protein  
F:1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>  
F:1-371/Domain: extracellular #status predicted <EXT>  
F:9-86/Domain: immunoglobulin homology <IM1>  
F:111-161/Domain: immunoglobulin homology #status atypical <IM2>  
F:191-274/Domain: immunoglobulin homology <IM3>  
F:296-347/Domain: immunoglobulin homology <IM4>  
F:372-395/Domain: transmembrane #status predicted <TM>  
F:396-432/Domain: intracellular #status predicted <INT>  
F:16-84,130-159,303-345/Disulfide bonds: #status predicted  
F:271,300/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.0%; Score 1876; DB 1; Length 432;  
Best Local Similarity 98.4%; Pred. No. 9, 1e-104;  
Matches 364; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY	27	KVLLGKGDVVELTCTASQKSIQPHWKNNOIKILGNQGSFLTKGPSKLNDRADRSRL	86
DB	2	KVLLGKGDVVELTCTASQKSIQPHWKNNOIKILGNQGSFLTKGPSKLNDRADRSRL	61
QY	87	WDQGNFLIKLKIETSDTYICEVDQKEVQLVFGLTANSDFLLQGSLLTTLTSP	146
DB	62	WDQGNFLIKLKIETSDTYICEVDQKEVQLVFGLTANSDFLLQGSLLTTLTSP	121
QY	147	PSSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTCTVLQNKVEFKIDIVLAFOK	206
DB	122	PSSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTCTVLQNKVEFKIDIVLAFOK	181
QY	207	ASSIVYKKEGEVSEFPLAFTVEKLTGSGELMWQAEASSSSKSWITFDLNKKEVSVKRV	266
DB	182	ASSIVYKKEGEVSEFPLAFTVEKLTGSGELMWQAEASSSSKSWITFDLNKKEVSVKRV	241
QY	267	TDDPKLQMGKPLHLTLTPQALPQYAGSGNLTALAEATGKLGHOVNLVVMRATQLOKXL	326
DB	242	TDDPKLQMGKPLHLTLTPQALPQYAGSGNLTALAEATGKLGHOVNLVVMRATQLOKXL	301
QY	327	TCEVWGPTSPKMLSLKLENKAKVSKREKPVVNLNPEAGMOCCLSDSGOVLLESNIKV	386
DB	302	TCEVWGPTSPKMLSLKLENKAKVSKREKAVVNLNPEAGMOCCLSDSGOVLLESNIKV	361
QY	387	LPTWSTPVEP	396
DB	362	LPTWSTPVEP	371

RESULT 3  
R/MQ74  
T-cell surface glycoprotein CD4 - rhesus macaque  
N/Alternate names: T-cell surface antigen T4/Len 3  
C/Species: Macaca mulatta (rhesus macaque)

C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999  
C/Accession: C32722  
R/Camerini, D.; Seed, B.  
Cell 60, 747-754, 1990  
A/Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the vi  
A/Reference number: A32722; MUID:90182664; PMID:2107024  
A/Accession: C32722  
A/Molecule type: mRNA  
A/Residues: 1-432 <CAM>  
A/Cross-references: GB:M31134  
C/Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells  
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
C/Keywords: duplication; glycoprotein; T-cell; transmembrane protein  
F:1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>  
F:1-371/Domain: extracellular #status predicted <EXT>  
F:9-86/Domain: immunoglobulin homology <IM1>  
F:111-161/Domain: immunoglobulin homology #status atypical <IM2>  
F:180-293/Domain: immunoglobulin homology <IM3>  
F:296-347/Domain: immunoglobulin homology <IM4>  
F:372-395/Domain: transmembrane #status predicted <TM>  
F:396-432/Domain: intracellular #status predicted <INT>  
F:16-84,130-159,303-345/Disulfide bonds: #status predicted  
F:271,300/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.9%; Score 1738; DB 1; Length 432;  
Best Local Similarity 90.5%; Pred. No. 1, 3e-95;  
Matches 335; Conservative 17; Mismatches 18; Indels 0; Gaps 0;

QY	27	KVLLGKGDVVELTCTASQKSIQPHWKNNOIKILGNQGSFLTKGPSKLNDRADRSRL	86
DB	2	KVLLGKGDVVELTCTASQKSIQPHWKNNOIKILGNQGSFLTKGPSKLNDRADRSRL	61
QY	87	WDQGNFLIKLKIETSDTYICEVDQKEVQLVFGLTANSDFLLQGSLLTTLTSP	146
DB	62	WDQGNFLIKLKIETSDTYICEVDQKEVQLVFGLTANSDFLLQGSLLTTLTSP	121
QY	147	PSSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTCTVLQNKVEFKIDIVLAFOK	206
DB	122	PSSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTCTVLQNKVEFKIDIVLAFOK	181
QY	207	ASSIVYKKEGEVSEFPLAFTVEKLTGSGELMWQAEASSSSKSWITFDLNKKEVSVKRV	266
DB	182	ASSIVYKKEGEVSEFPLAFTVEKLTGSGELMWQAEASSSSKSWITFDLNKKEVSVKRV	241
QY	267	TDDPKLQMGKPLHLTLTPQALPQYAGSGNLTALAEATGKLGHOVNLVVMRATQLOKXL	326
DB	242	TDDPKLQMGKPLHLTLTPQALPQYAGSGNLTALAEATGKLGHOVNLVVMRATQLOKXL	301
QY	327	TCEVWGPTSPKMLSLKLENKAKVSKREKPVVNLNPEAGMOCCLSDSGOVLLESNIKV	386
DB	302	TCEVWGPTSPKMLSLKLENKATVSKQAKAVVNLNPEAGMOCCLSDSGOVLLESNIKV	361
QY	387	LPTWSTPVEP	396
DB	362	LPTWSTPVEP	371

RESULT 4  
GHU  
IG gamma-1 chain C region - human  
C/Species: Homo sapiens (man)  
C/Date: 31-Jan-1981 #sequence\_revision 18-Aug-1982 #text\_change 16-Jul-1999  
C/Accession: A93433; S36861; S33887; B90563; A90564; B91668; A91723; A02146  
R/Elison, J.W.; Berzon, B.J.; Hood, L.E.  
Nucleic Acids Res. 10, 4071-4079, 1982  
A/Title: The nucleotide sequence of a human immunoglobulin C-gamma1 gene.  
A/Reference number: A93433; MUID:82274238; PMID:6287432  
A/Accession: A93433  
A/Molecule type: DNA  
A/Residues: 1-330 <ELL>  
A/Cross-references: EMBL:217370  
A/Note: this sequence has the Gln(17) allelic marker, 97-Lye, and the Gln(1) markers,  
A:Note: Lys-330 is removed after translation



R.Harris, L.J.  
submitted to the EMBL Data Library, October 1992  
A:Reference number: S33904  
A:Accession: S36861  
A:Molecule type: DNA  
A:Residues: 2-330 <HAR>  
A:Cross-references: EMBL:Z17370  
R:Yakubashvili, N.; Ueda, S.; Odate, M.; Nikaido, T.; Nakai, S.; Honjo, T.  
Cell 29, 671-679, 1982  
A:Title: Structure of human immunoglobulin gamma genes: implications for evolution of a  
A:Reference number: S33887; MUID:83001943; PMID:6811139  
A:Accession: S33887  
A:Molecule type: DNA  
A:Residues: 88-113/235-330 <TAK>  
A:Cross-references: EMBL:Z17370  
R:Cunningham, B.A.; Rutishauser, U.; Gall, W.E.; Gottlieb, P.D.; Waxdal, M.J.; Edelman,  
Biochemistry 9, 3161-3170, 1970  
A:Title: The covalent structure of a human gammag-1 immunoglobulin. VII. Amino acid sequen  
A:Reference number: A90563; MUID:71064024; PMID:5489771  
A:Contents: myeloma protein Eu  
A:Accession: B90563  
A:Molecule type: protein  
A:Residues: 1-96, 'R', 98-135 <GUN>  
A:Note: This sequence has the G1m(3) marker, 97-Arg  
R:Rutishauser, U.; Cunningham, B.A.; Bennett, C.; Konigsberg, W.H.; Edelman, G.M.  
Biochemistry 9, 3171-3181, 1970  
A:Title: The covalent structure of a human gammag-1 immunoglobulin. VIII. Amino acid sequen  
A:Reference number: A90564; MUID:71064025; PMID:5530842  
A:Contents: Eu  
A:Accession: A90564  
A:Molecule type: protein  
A:Residues: 136-154, 'Q', 156-165, 'Q', 167-176, 'Q', 178-194, 'N', 196-197, 'D', 199-238, 'E', 240,  
A:Note: This sequence has the G1m(non-1) markers, 239-Glu and 241-Met  
R:Postnig, H.; Hilschmann, N.  
Hoppe-Seyler's Z. Physiol. Chem. 357, 1571-1604, 1976  
A:Title: Die Primerstruktur eines monoklonalen IgG1-Immunglobulins (Myelomprotein Nie),  
Igen Primerstruktur.  
A:Reference number: A91668; MUID:77070269; PMID:826475  
A:Contents: myeloma protein Nie  
A:Accession: B91668  
A:Molecule type: protein  
A:Residues: 1-34, 'Q', 36-96, 'K', 98-115, 'Q', 117-197, 'D', 199-238, 'D', 240, 'L', 242-268, 'E', 27  
A:Note: This sequence has the G1m(17) and G1m(1) markers  
R:Schmidt, W.E.; Jung, H.D.; Palm, W.; Hilschmann, N.  
Hoppe-Seyler's Z. Physiol. Chem. 364, 713-747, 1983  
A:Title: Die Primerstruktur des kristallisierten monoklonalen Immunglobulins IgG1 KOL  
A:Reference number: A91723; MUID:83289131; PMID:6884994  
A:Contents: myeloma protein KOL; disulfide bonds  
A:Accession: A91723  
A:Molecule type: protein  
A:Residues: 1-96, 'R', 98-197, 'D', 199-238, 'E', 240, 'M', 242-266, 'D', 268-271, 'D', 273-330 <SCH  
R:Gall, W.E.; Edelman, G.M.  
Biochemistry 9, 3188-3196, 1970  
A:Title: The covalent structure of a human gammag-1 immunoglobulin. X. Intrachain disulfide  
A:Reference number: A90565; MUID:71064027; PMID:4923144  
A:Contents: annotation; disulfide bonds  
R:Diexler, L.; Schwarz, J.; Reichel, W.; Hilschmann, N.  
Hoppe-Seyler's Z. Physiol. Chem. 357, 1515-1540, 1976  
A:Title: Rule of antibody structure. The primary structure of monoclonal IgG1 immunoglob  
enbromide cleavage products, and the disulfide bridges.  
A:Reference number: A91667; MUID:77070267; PMID:1002129  
A:Contents: annotation; disulfide bonds  
C:Genetics:  
A:Gene: GDB:IGHG1  
A:Cross-references: GDB:120085; OMTM:147100  
A:Map position: 14q32.33-14q32.33  
A:Introns: 99/1; 114/1; 224/1  
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)  
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin  
F:20-85/Domain: immunoglobulin homology <IMI>

F:137-206/Domain: immunoglobulin homology <IM2>  
F:243-310/Domain: immunoglobulin homology <IM3>  
F:27-83,144-204,250-308/Disulfide bonds: #status experimental  
F:103/Disulfide bonds: interchain (to light chain) #status experimental  
F:109,112/Disulfide bonds: interchain (to heavy chain) #status experimental  
F:180/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 37.1%; Score 1266; DB 1; Length 330;  
Best Local Similarity 96.7%; Pred. No. 86-68;  
Matches 236; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

QY 382 SNIKVLPFTWSTPVPKSCDTHTCPCPAPBELLGSPVFLPPPKDTLMISRPEVTCV 441  
|||  
DB 90 SNTKCV---DKKVEPKSCDTHTCPCPAPBELLGSPVFLPPPKDTLMISRPEVTCV 145  
QY 442 VVDVSHDEPVEKFMWYDGVVHNAKTPREBOVNSYRVSVLTVLHODMLNGEKYCK 501  
|||  
DB 146 VVDVSHDEPVEKFMWYDGVVHNAKTPREBOVNSYRVSVLTVLHODMLNGEKYCK 205  
QY 502 VSNKALPAPLEKTIKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFYPSDIAVEME 561  
|||  
DB 206 VSNKALPAPLEKTIKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFYPSDIAVEME 265  
QY 562 SNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRWQGNVFCGSVMEALHNHYTQKSLS 621  
|||  
DB 266 SNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRWQGNVFCGSVMEALHNHYTQKSLS 325  
QY 622 LSPG 625  
|||  
DB 326 LSPG 329

## RESULT 5

S31866  
Ig gamma-1 chain C region - synthetic

C:Species: synthetic  
A:Note: Homo sapiens (man) gene engineered and expressed in Escherichia coli  
C:Date: 06-Jan-1995 #sequence\_revision 17-Mar-1997 #text\_change 19-May-2000  
C:Accession: S31866  
R:Filipula, D.  
submitted to the EMBL Data Library, February 1993  
A:Description: Screening method for protein-protein interactions of cloned gene products.  
A:Reference number: S31866  
A:Accession: S31866  
A:Molecule type: mRNA  
A:Residues: 1-255 <FIL>  
A:Cross-references: EMBL:X70421; NID:G33068; PIDN:CAA49866.1; PID:G33069  
C:Keywords: immunoglobulin  
F:1-22/Region: Escherichia coli outer membrane protein A precursor  
F:23-255/Region: human Ig gamma-1 chain C region

Query Match 36.7%; Score 1254; DB 4; Length 255;  
Best Local Similarity 99.6%; Pred. No. 36-67; 1; Indels 0; Gaps 0;  
Matches 231; Conservative 0; Mismatches 1;

QY 394 VEPKSCDTHTCPCPAPBELLGSPVFLPPPKDTLMISRPEVTCVVDVSHDEPEVK 453  
|||  
DB 23 VESKSCDTHTCPCPAPBELLGSPVFLPPPKDTLMISRPEVTCVVDVSHDEPEVK 82  
QY 454 FNNYVDGVVHNAKTPREBOVNSYRVSVLTVLHODMLNGEKYCKKVSNAKLPAPLEK 513  
|||  
DB 83 FNNYVDGVVHNAKTPREBOVNSYRVSVLTVLHODMLNGEKYCKKVSNAKLPAPLEK 142  
QY 514 TISKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFYPSDIAVEMSGQPENNYKTT 573  
|||  
DB 143 TISKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFYPSDIAVEMSGQPENNYKTT 202  
QY 574 PVLVDSDGSFPLYSKLTVDKSRWQGNVFCGSVMEALHNHYTQKSLSLSPG 625  
|||  
DB 203 PVLVDSDGSFPLYSKLTVDKSRWQGNVFCGSVMEALHNHYTQKSLSLSPG 254

## RESULT 6



S69339  
 Ig heavy chain V region precursor - human  
 C:Species: Homo sapiens (man)  
 C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 01-Dec-2000  
 C:Accession: S69339; S72664  
 R:Khanlich, A.A.; Aucouturier, P.; Preud'homme, J.L.; Cogné, M.  
 Eur. J. Biochem. 229, 54-60, 1995  
 A:Title: Structure of abnormal heavy chains in human heavy-chain-deposition disease.  
 A:Reference number: S69339; MUID:95262687; PMID:7744049  
 A:Accession: S69339  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-374 <KHA>  
 A:Cross-references: EMBL:X81695  
 R:Khanlich, A.A.  
 submitted to the EMBL Data Library, September 1994  
 A:Reference number: S72664  
 A:Accession: S72664  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-140 'C', 142-374 <KH2>  
 A:Cross-references: EMBL:X81695  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology

Query Match 36.7%; Score 1252.5; DB 2; Length 374;  
 Best Local Similarity 99.1%; Pred. No. 5, 9e-67;  
 Matches 229; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 395 EPKSCDKHTCPCPAPPELLGSPVFLPPPKDITMISRTPEVTCVVVDVSHEDPEVKF 454  
 DB 143 EPKSCDKHTCPCPAPPELLGSPVFLPPPKDITMISRTPEVTCVVVDVSHEDPEVKF 202

QY 455 NMYVDGVEVHNKTKPREQYNSTYRVSVLTGLHQMNLNGEKYCKSNKALPAPIEKT 514  
 DB 203 NMYVDGVEVHNKTKPREQYNSTYRVSVLTGLHQMNLNGEKYCKSNKALPAPIEKT 262

QY 515 ISKAKGQPREPVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTP 574  
 DB 263 ISKAKGQPREPVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTP 322

QY 575 PVLDSGSEFPLYSKLTVDKSRWQGNVFCSVMEALHNHYTQKSLSLSPG 625  
 DB 323 PVLDSGSEFPLYSKLTVDKSRWQGNVFCSVMEALHNHYTQKSLSLSPG 373

RESULT 7  
 PT0207  
 Ig gamma chain C region - chimpanzee  
 C:Species: Pan troglodytes (chimpanzee)  
 C>Date: 23-Nov-1991 #sequence\_revision 23-Nov-1991 #text\_change 16-Jul-1999  
 C:Accession: PT0207  
 R:Brilich, P.H.; Moustafa, Z.A.; Oestberg, L.  
 Mol. Immunol. 28, 319-322, 1991  
 A:Title: Nucleotide sequence of chimpanzee Cc and hinge regions.  
 A:Reference number: PT0207; MUID:91287716; PMID:2062315  
 A:Accession: PT0207  
 A:Molecule type: mRNA  
 A:Residues: 1-234 <BHR>  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: immunoglobulin  
 F:48-117/Domain: immunoglobulin homology <IMM>

Query Match 35.7%; Score 1218; DB 2; Length 234;  
 Best Local Similarity 95.4%; Pred. No. 3, 6e-65;  
 Matches 227; Conservative 1; Mismatches 6; Indels 4; Gaps 1;

QY 382 SNIKVLPWTSTPEPKSCDKHTCPCPAPPELLGSPVFLPPPKDITMISRTPEVTCV 441  
 DB 1 SNIKVLPWTSTPEPKSCDKHTCPCPAPPELLGSPVFLPPPKDITMISRTPEVTCV 56

QY 442 VVDVSHEDPEVKFNMYVDGVEVHNKTKPREQYNSTYRVSVLTGLHQMNLNGEKYCK 501  
 DB 57 VVDVSHEDPEVKFNMYVDGVEVHNKTKPREQYNSTYRVSVLTGLHQMNLNGEKYCK 116

QY 502 VSNKALPAPIEKTISKAKGQPREPVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVW 561  
 DB 117 VSNKALPAPIEKTISKAKGQPREPVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVW 176

QY 562 SNGQPENNYKTPPVLDSDGSEFPLYSKLTVDKSRWQGNVFCSVMEALHNHYTQKS 619  
 DB 177 SNGQPENNYKTPPVLDSDGSEFPLYSKLTVDKSRWQGNVFCSVMEALHNHYTQKS 234

RESULT 8  
 A23511  
 Ig gamma-3 chain C region (allotype G3m(b)) - human  
 C:Species: Homo sapiens (man)  
 C>Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 23-Jul-1999  
 C:Accession: A23511  
 R:Huck, S.; Fort, P.; Crawford, D.H.; Lefranc, M.P.; Lefranc, G.  
 Nucleic Acids Res. 14, 1779-1789, 1986  
 A:Title: Sequence of a human immunoglobulin gamma 3 heavy chain constant region gene: cc  
 A:Reference number: A23511; MUID:86148507; PMID:3081877  
 A:Accession: A23511  
 A:Molecule type: DNA  
 A:Residues: 1-377 <HUC>  
 A:Cross-references: GB:X03604; GB:M12958; NID:933070; PIDN:CAA27268.1; PID:9577056  
 A:Gene: IGHG3  
 A:Cross-references: GDB:119339; OMIM:147120  
 A:Map position: 14q32.33-14q32.33  
 A:Intron: 98/3; 115/3; 130/3; 145/3; 160/3; 270/3  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: immunoglobulin  
 F:20-85/Domain: immunoglobulin homology <IMM>

Query Match 34.3%; Score 1171; DB 2; Length 377;  
 Best Local Similarity 92.6%; Pred. No. 4e-62;  
 Matches 214; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 395 EPKSCDKHTCPCPAPPELLGSPVFLPPPKDITMISRTPEVTCVVVDVSHEDPEVKF 454  
 DB 146 EPKSCDKHTCPCPAPPELLGSPVFLPPPKDITMISRTPEVTCVVVDVSHEDPEVKF 205

QY 455 NMYVDGVEVHNKTKPREQYNSTYRVSVLTGLHQMNLNGEKYCKSNKALPAPIEKT 514  
 DB 206 NMYVDGVEVHNKTKPREQYNSTYRVSVLTGLHQMNLNGEKYCKSNKALPAPIEKT 265

QY 515 ISKAKGQPREPVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTP 574  
 DB 266 ISKAKGQPREPVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTP 325

QY 575 PVLDSGSEFPLYSKLTVDKSRWQGNVFCSVMEALHNHYTQKSLSLSPG 625  
 DB 326 PVLDSGSEFPLYSKLTVDKSRWQGNVFCSVMEALHNHYTQKSLSLSPG 376

RESULT 9  
 A60764  
 Ig gamma-3 chain C region, form LAT - human  
 C:Species: Homo sapiens (man)  
 C>Date: 14-May-1993 #sequence\_revision 14-May-1993 #text\_change 16-Jul-1999  
 C:Accession: A60764  
 R:Huck, S.; Lefranc, G.; Lefranc, M.P.  
 Immunogenetics 30, 250-257, 1989  
 A:Title: A human immunoglobulin IGHG3 allele (Gmb0, b1, c3, c5, u) with an IGHG4 convert  
 A:Reference number: A60764; MUID:90007613; PMID:2571587  
 A:Accession: A60764  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-377 <HUC>  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: immunoglobulin  
 F:20-85/Domain: immunoglobulin homology <IMM>

Query Match 34.2%; Score 1169; DB 2; Length 377;



Best Local Similarity 92.6%; Pred. No. 5.2e-62;  
Matches 214; Conservative 8; Mismatches 9; Indels 0; Gaps 0;  
QY 395 EPKSCDTHNCPCPAPELLGSPSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 454  
D 146 EPKSCDTHNCPCPAPELLGSPSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 205  
QY 455 NWYVDGVEVNAKTKPREEQYNSTYRVSVLTVLHQMVLNGKEYKCKVSKALPAPIEKT 514  
D 206 KMYVDGVEVNAKTKPREEQYNSTYRVSVLTVLHQMVLNGKEYKCKVSKALPAPIEKT 265  
QY 515 ISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTP 574  
D 266 ISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTP 325  
QY 575 PVLSDSGSFPLYSKLTVDKSRMOQGNVFCSCVMHEALHNHYTOKSLSPG 625  
D 326 PVLSDSGSFPLYSKLTVDKSRMOQGNVFCSCVMHEALHNHYTOKSLSPG 376

## RESULT 10

G3HUW1

Ig gamma-3 heavy chain disease proteins - human

C/Species: Homo sapiens (man)

C/Date: 31-Dec-1979 #sequence\_revision 23-Oct-1981 #text change 16-Jul-1999

C/Accession: A90442; A92219; A90198; A93915; A02149

R/Frangiione, B.; Rosenwasser, E.; Prellil, F.; Franklin, E.C.

Biochemistry 19, 4304-4308, 1980

A/Title: Primary structure of human gamma3 immunoglobulin deletion mutant: gamma3 heavy-

A/Reference number: A90442; MUID:81021548; PMID:6774747

A/Contents: heavy chain disease protein W15

A/Accession: A90442

A/Molecule type: protein

A/Residues: 1-289 &lt;FRA&gt;

A/Note: the molecule is a dimer linked by 12 disulfide bonds; it has an extra interchain

A/Note: this protein lacks most of the V region and all of the CH1 region. Residue 12 co

R/Michaelson, T.E.; Frangiione, B.; Franklin, E.C.

J. Biol. Chem. 252, 883-889, 1977

A/Title: Primary structure of the 'hinge' region of human IgG3. Probable quadruplication

A/Reference number: A92219; MUID:77110561; PMID:402363

A/Contents: normal gamma-3 chains, sequence corresponding to residues 12-97 of protein W

A/Accession: A92219

A/Molecule type: protein

A/Residues: 12-97 &lt;MIC&gt;

A/Note: the hinge region in gamma-3 chains is about four times as long as in other gamma

A/Note: cysteines at positions 24, 27, 33, 39, 42, 48, 54, 57, 63, 69, and 72 form inter

R/Wolfsenstein-Todel, C.; Frangiione, B.; Prellil, F.; Franklin, E.C.

Biochem. Biophys. Res. Commun. 71, 907-914, 1976

A/Title: The amino acid sequence of "heavy chain disease" protein ZUC. Structure of the

A/Reference number: A90198; MUID:77021516; PMID:823945

A/Contents: heavy chain disease protein ZUC, partial sequence corresponding to residues

A/Accession: A90198

A/Molecule type: protein

A/Residues: 59-125, 'EB', 128-226, 228-289 &lt;WOL&gt;

A/Note: this protein lacks most of the V region, all of the CH1 region, and part of the

R/Alexander, A.; Steimetz, M.; Barriatua, D.; Frangiione, B.; Franklin, E.C.; Hood, L.;

Proc. Natl. Acad. Sci. U.S.A. 79, 3260-3264, 1982

A/Title: gamma heavy chain disease in man: cDNA sequence supports partial gene deletion

A/Reference number: A93915; MUID:82247835; PMID:6808505

A/Contents: heavy chain disease protein Omh

A/Accession: A93915

A/Molecule type: mRNA

A/Residues: 12-70, 72-114, 116-125, 'E', 127-133, 'L', 135-136, 'E', 138, 'Y', 140-154, 'D', 156-157

A/Note: a carboxyl-terminal Lys is removed posttranslationally

A/Note: this sequence may represent an allelic form or another gamma chain subclass

C/Comment: The heavy chain disease protein W15 is shown.

C/Genetics:

A/Genes: GDB:IGHG3

A/Cross-references: GDB:119339; OMIM:147120

A/Map position: 14q32.33-14q32.33

C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: duplication; glycoprotein; immunoglobulin; pyroglutamic acid  
F:203-270/Domain: immunoglobulin homology <IMM>  
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
F:6,140/Birding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 33.7%; Score 1151; DB 1; Length 289;  
Best Local Similarity 90.5%; Pred. No. 4.3e-61;  
Matches 209; Conservative 13; Mismatches 9; Indels 0; Gaps 0;  
QY 395 EPKSCDTHNCPCPAPELLGSPSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 454  
D 59 EPKSCDTHNCPCPAPELLGSPSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 118  
QY 455 NWYVDGVEVNAKTKPREEQYNSTYRVSVLTVLHQMVLNGKEYKCKVSKALPAPIEKT 514  
D 119 KMYVDGVEVNAKTKPREEQYNSTYRVSVLTVLHQMVLNGKEYKCKVSKALPAPIEKT 178  
QY 515 ISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTP 574  
D 179 ISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNGQPENNYKTTP 238

QY 575 PVLSDSGSFPLYSKLTVDKSRMOQGNVFCSCVMHEALHNHYTOKSLSPG 625  
D 239 PVLSDSGSFPLYSKLTVDKSRMOQGNVFCSCVMHEALHNHYTOKSLSPG 289

## RESULT 11

G2HU

Ig gamma-2 chain C region - human

C/Species: Homo sapiens (man)

C/Date: 30-Apr-1981 #sequence\_revision 13-Jun-1983 #text change 21-Jul-2000

C/Accession: A93906; A92809; A90752; A93132; A02148

R/Elliott, J.; Hood, L.

Proc. Natl. Acad. Sci. U.S.A. 79, 1984-1988, 1982

A/Title: Linkage and sequence homology of two human immunoglobulin gamma heavy chain con

A/Reference number: A93906; MUID:82197621; PMID:6804948

A/Accession: A93906

A/Molecule type: DNA

A/Residues: 1-326 &lt;ELL&gt;

A/Note: references: GB:Y00554; GB:Y00230; NID:G32759; PIDN:CA85438.1; PID:G606056

A/Note: Lys-326 is probably removed posttranslationally

R/Wang, A.C.; Tung, E.; Fudenberg, H.H.

J. Immunol. 125, 1048-1054, 1980

A/Title: The primary structure of a human IgG2 heavy chain: genetic, evolutionary, and f

A/Reference number: A92809; MUID:81007873; PMID:6774012

A/Contents: myeloma protein T11

A/Accession: A92809

A/Molecule type: protein

A/Residues: 1-19, 'Q', 21-57, 'Z', 59, 'A', 61-193, 'D', 195-325 &lt;WAN&gt;

A/Note: Trp-156 is at or near the complement-binding site

R/Connell, G.E.; Parr, D.M.; Hofmann, T.

Can. J. Biochem. 57, 758-767, 1979

A/Title: The amino acid sequences of the three heavy chain constant region domains of a t

A/Reference number: A90752; MUID:80001357; PMID:113060

A/Contents: myeloma protein Z1e

A/Accession: A90752

A/Molecule type: protein

A/Residues: 1-24, 'E', 26-57, 'EV', 60-85, 132-171, 'ZZZ', 175, 'B', 177-193, 'D', 195-196, 'Q', 198-2

A/Note: this sequence has since been revised

R/Hofmann, T.; Parr, D.M.

Mol. Immunol. 16, 923-925, 1979

A/Title: A note on the amino acid sequence of residues 381-391 of human immunoglobulin g

A/Reference number: A93132; MUID:80114419; PMID:118920

A/Contents: Z1e

A/Accession: A93132

A/Molecule type: protein

A/Residues: 238-275 &lt;HOF&gt;

R/Hofmann, T.; Parr, D.M.

submitted to the Atlas, March 1980

A/Reference number: A94591

A/Contents: annotation; Z1e, revisions to residues 25, 59, 60, and 264-268

A/Note: the revised sequence differs from that shown in having 60-Ala and in the amidat

ned



R.Milstein, C.; Frangione, B.  
 Biochem. J. 121, 217-225, 1971  
 A>Title: Disulfide bridges of the heavy chain of human immunoglobulin G2.  
 A:Reference number: A90253; MUID:72033500; PMID:4940472  
 A:Contents: annotation; myeloma protein S<sub>μ</sub>, disulfide bonds  
 R:Frangione, B.; Milstein, C.; Pink, J.R.L.  
 Nature 221, 145-148, 1969  
 A>Title: Structural studies of immunoglobulin G.  
 A:Reference number: A9157; MUID:69064124; PMID:5182707  
 A:Contents: annotation; S<sub>μ</sub>, disulfide bonds  
 C:Genetics:  
 A:Gene: GDB:IGHG2  
 A:Cross-references: GDB:119338; OMIM:147110  
 A:Map position: 14q32.33-14q32.33  
 C:Map: position: An immunoglobulin heterotrimer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin  
 F:120-85/Domain: immunoglobulin homology <IM1>  
 F:133-202/Domain: immunoglobulin homology <IM2>  
 F:239-306/Domain: immunoglobulin homology <IM3>  
 F:14/Disulfide bonds: interchain (to light chain) #status experimental  
 F:27-83,140-200,246-304/Disulfide bonds: #status experimental  
 F:102,103,106,109/Disulfide bonds: interchain (to heavy chain) #status experimental  
 F:176/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 33.7%; Score 1150; DB 1; Length 326;  
 Best Local Similarity 88.9%; Pred. No. 5,8e-61;  
 Matches 217; Conservative 9; Mismatches 10; Indels 8; Gaps 3;

QY 382 SNIKVLTWSTVEPEKSCDKHTTCCPAPELLGSGVLPFPKPKDTLMISRTPEVTCV 441  
 DB 90 SNTKVDKT-----VERKCCVE---CPCCPAP-P-VAGPSVLPFPKPKDTLMISRTPEVTCV 141

QY 442 VVDVSHEDPEVKFNRYVDGEVYNAKTKPREQYNSTRVSVLTVLHDMINGEKYCK 501  
 DB 142 VVDVSHEDPEVKFNRYVDGEVYNAKTKPREQYNSTRVSVLTVLHDMINGEKYCK 201

QY 502 VSNKLLPAPIEKTISAKKQPREPOVYTLPPSRDELTKNOVSLTCLVNGFYPSDIAVEWE 561  
 DB 202 VSNKGLPAPIEKTISAKKQPREPOVYTLPPSRDELTKNOVSLTCLVNGFYPSDIAVEWE 261

QY 562 SNGQENNYKTPPVLDSGSEFLYSKLTVDKSRMQQGNVFCSSVMHEALNHYTQKSL 621  
 DB 262 SNGQENNYKTPPVLDSGSEFLYSKLTVDKSRMQQGNVFCSSVMHEALNHYTQKSL 321

QY 622 LSPG 625  
 DB 322 LSPG 325

RESULT 12  
 A66254  
 CD4 precursor - rabbit  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jan-2000  
 C:Accession: A46254  
 R:Haeghe, B.F.; Savadkoe, S.; Brown, T.J.; Lee, K.; Recker, D.P.; Kindt, T.J.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 7963-7967, 1992  
 A>Title: CD4 and its role in infection of rabbit cell lines by human immunodeficiency virus  
 A:Reference number: A46254; MUID:92390370; PMID:1518821  
 A:Accession: A46254  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-459 <HAG>  
 A:Cross-references: GDB:M92840; NID:9164871; PID:AAA1198.1; PID:9164872  
 A>Note: sequence extracted from NCBI backbone (NCBIN:112732, NCBI:112733)  
 C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
 F:322-372/Domain: immunoglobulin homology <IM1>

Query Match 33.7%; Score 1149; DB 2; Length 459;  
 Best Local Similarity 56.9%; Pred. No. 1e-60;  
 Matches 243; Conservative 66; Mismatches 98; Indels 20; Gaps 6;

QY 1 MNRGVPFRLTLVLTALLPATQGNKVKLGKGPVETLTCTASOKKSIOPHMKNNOIK 60  
 DB 1 MNRGVPFRLTLVLTALLPATQGNKVKLGKGPVETLTCTASOKKSIOPHMKNNOIK 60

QY 61 ILNGG----GFLTKGPKLDRADSRSLDQGNFPLIINKLIEDSYICEVEDQKE 116  
 DB 61 ILNGGSSSSPFLKGNPLSNRVSKKMMMDQGSFPLVIDLRNDDSGTYICEVEDQKE 120

QY 117 EVQLLVFGLTANSDPHLLQGSLTITLESPPSSSVSCSPRGNIGGKTLISVQLEL 176  
 DB 121 EVELLVFRLTANPNRLRHGSLTITLESPPSSSVSCSPRGNIGGKTLISVQLEL 180

QY 177 QDSGWTCTV-LQNGKVEPKIDIVLAFAKASSIVYKKEGQVFFSPPLAFVTKLGS 235  
 DB 181 QDSGWSCHLSFPQDNKLELKIIVLFPKASATVYKKEGQVFFSPPLAFVTKLGS 238

QY 236 GELMWQAERASSSKSWITFDLKNKESVYKATVOTDPKLPLHLTPQALPOYAGSG 295  
 DB 239 GELMWQVAGASAGSWSFSLDRKVSQKILPLDKIQMSGLPLSLTPQALHRYAGSG 298

QY 296 NITLALFKTKLHOEVNLVVMRATQLOKNTCEYWGFTSPKMLSLKLENKAKVSKRE 355  
 DB 299 NITSLTLD--KSKLHQVSLVMLKVTQVKNKLTCEVLTGPIIDPKMLSLKLEDOEAKVS-TQ 355

QY 356 KPVWLVNPEAGMWQCLSDSGOVLLESNIKVLPTWSTVEPEKSCDKHTTCCPAPPELLG 415  
 DB 356 KPVWLVNPEAGMWQCLSDSGOVLLESNIKVLPTWSTVEPEKSCDKHTTCCPAPPELLG 405

QY 416 GPSVFLF 422  
 DB 406 TAGLVLF 412

RESULT 13  
 G4HU  
 Ig gamma-4 chain C region - human  
 C:Species: Homo sapiens (man)  
 C:Date: 02-Apr-1982 #sequence\_revision 02-Apr-1982 #text\_change 16-Jul-1999  
 C:Accession: A90933; A90249; A02150  
 R:Ellison, J.; Buxbaum, J.; Hood, L.  
 DNA 1, 11-18, 1981  
 A>Title: Nucleotide sequence of a human immunoglobulin C-gamma4 gene.  
 A:Reference number: A90933; MUID:83157104; PMID:6299662  
 A:Accession: A90933  
 A:Molecule type: DNA  
 A:Residues: 1-327 <ELU>  
 A>Note: the sequence was determined from the germ-line gene  
 R:Pink, J.R.L.; Buttery, S.H.; De Vries, G.M.; Milstein, C.  
 Biochem. J. 117, 33-47, 1970  
 A>Title: Human immunoglobulin subclones. Partial amino acid sequence of the constant r  
 A:Reference number: A90249; MUID:70207560; PMID:4192699  
 A:Accession: A90249  
 A:Molecule type: protein  
 A:Residues: 1-30;81-326 <PIN>  
 C:Genetics:  
 A:Gene: GDB:IGHG4  
 A:Cross-references: GDB:119340; OMIM:147130  
 A:Map position: 14q32.33-14q32.33  
 A:Introns: 99/1; 111/1; 221/1  
 C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin  
 F:20-85/Domain: immunoglobulin homology <IM1>  
 F:99-110/Region: hinge  
 F:134-203/Domain: immunoglobulin homology <IM2>  
 F:240-307/Domain: immunoglobulin homology <IM3>  
 F:14/Disulfide bonds: interchain (to light chain) #status experimental  
 F:27-83,141-201,247-305/Disulfide bonds: #status predicted  
 F:106,109/Disulfide bonds: interchain (to heavy chain) #status experimental  
 F:177/Binding site: carbohydrate (Asn) (covalent) #status predicted







A/Molecule type: mRNA  
A/Residues: 1-457 <TOU>  
A/Cross-references: GB:M13816; NID:G192070; PIDN:AAA7267.1; PID:G309112  
R.Littman, D.R.; Gettner, S.N.  
Nature 325, 453-455, 1987  
A/Title: Unusual intron in the immunoglobulin domain of the newly isolated murine CD4 (L  
A/Reference number: A26038; MUID:87115821; PMID:3027575  
A/Accession: A26038  
A/Molecule type: mRNA  
A/Residues: 1-457 <LIT>  
A/Cross-references: GB:X04836; NID:G50353; PIDN:CAA28539.1; PID:G50354  
R.Gorman, S.D.; Tourville, B.; Parne, J.R.  
Proc. Natl. Acad. Sci. U.S.A. 84, 7644-7648, 1987  
A/Title: Structure of the mouse gene encoding CD4 and an unusual transcript in brain.  
A/Reference number: A39893; MUID:88041159; PMID:2823269  
A/Accession: A39893  
A/Molecule type: DNA  
A/Residues: 1-25, 'E', 27-457 <GOR>  
A/Cross-references: GB:M17080; GB:J03003; NID:G192515; PIDN:AAA37402.1; PID:G387124  
R.Maddon, P.J.; Molineux, S.M.; Maddon, D.E.; Zimmerman, K.A.; Godfrey, M.; Alt, F.W.;  
Proc. Natl. Acad. Sci. U.S.A. 84, 9155-9159, 1987  
A/Title: Structure and expression of the human and mouse T4 genes.  
A/Reference number: A39955; MUID:88097446; PMID:3501122  
A/Accession: A39955  
A/Status: nucleic acid sequence not shown; not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 25-457 <MAD>  
A/Note: the cited GenBank accession number, J03564, is not in release 101.0  
R.Parne, J.R.; Hunkapiller, T.  
Immunol. Rev. 100, 109-127, 1987  
A/Title: L3T4 and the immunoglobulin gene superfamily: New relationships between the imm  
A/Reference number: 154564; MUID:88152875; PMID:3326618  
A/Accession: 154564  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-457 <RES>  
A/Cross-references: GB:M36850; NID:G198670; PIDN:AAA39401.1; PID:G198671  
A/Accession: 169018  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 208-318 <RES2>  
A/Cross-references: GB:M36851; NID:G198672; PIDN:AAA39402.1; PID:G554183  
R.Clason, B.J.; Teagarden, J.; Kitzbaum, L.; Maddox, J.; Mackay, C.R.; Brandon, M.; M  
Immunogenetics 23, 129-132, 1986  
A/Title: The L3T4 antigen in mouse and the sheep equivalent are immunoglobulin-like.  
A/Reference number: A47642; MUID:86166694; PMID:3082751  
A/Accession: A47642  
A/Molecule type: protein  
A/Residues: 27-43 <CLA>  
A/Note: This protein is expressed on most thymocytes, on a subset of mature T-cells  
C/Genetics: 18/1; 74/1; 128/1; 207/1; 319/1; 386/1; 425/3; 448/2  
A/Introns: 18/1; 74/1; 128/1; 207/1; 319/1; 386/1; 425/3; 448/2  
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
C/Keywords: alternative initiators; duplication; glycoprotein; T-cell; transmembrane pro  
F.1-26/Domain: signal sequence #status predicted <Sig>  
F.1-27-457/Domain: T-cell surface glycoprotein CD4 #status experimental <MAT>  
F.35-114/Domain: immunoglobulin homology <IM1>  
F.119-190/Domain: immunoglobulin homology #status atypical <IM2>  
F.220-301/Domain: immunoglobulin homology <IM3>  
F.241-457/Domain: CD4, brain-specific short form #status predicted <BRA>  
F.321-372/Domain: immunoglobulin homology <IM4>  
F.395-419/Domain: transmembrane #status predicted <TM>  
F.420-457/Domain: intracellular #status predicted <INT>  
F.442-112,159-188,328-370/Denulfide bonds: #status predicted  
F.187,298,323,392/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 29.1%; Score 993; DB 1; Length 457;  
Best Local Similarity 53.8%; Pred. No. 1.7e-51;  
Matches 211; Conservative 62; Mismatches 109; Indels 10; Gaps 7;

QY 1 MNRGVFRH-LLLVQLALLPATOQNKVLGKGGDTVELTCTASQKSIQPHMNKSNQI 59  
DB 1 MCRATSLRLLLLQLLSQGLAVTQKTLVLGKGGSAELPCSSQKTLTFTWKFSDQR 60

QY 60 KILGNQ-SPLTKG--PSKLNDRADRSRLMDQGNFPIIKNLKLEDSDTYICEVEDQKE 116  
DB 61 KILGQHGKGLVLRGSSPSQF-DRFDSKKGAMEKSFPIIKNLKLEDSQTIICELENKE 119  
QY 117 EVQLVFLGRLNSDTHLQSGSLTLTLLES-PPGSSPVQCSPPKKNIOGQKTLSSVQLE 175  
DB 120 EVELWVFKYTPSPGSLSLQGGSLTLTLDSNSKVSPLRECKHKKKGVSGSVLSMSNLR 179  
QY 176 LODSGTWCTVLQNKVPEKIDIVLAFQKASIVYKKEGQVFPPLAFTVKLTGS 235  
DB 180 VQDSFPMKCTVTLDDCKMFGMTLSVLDFQSTALRAYSEGSAPFPPLNFAEB--NGW 237  
QY 236 GELTMQARASSSKSMTFEDDKNKVSRYRTQDPLQMGKLPILTLTPALPOLYAGSG 295  
DB 238 GELMKAEKDSFPQWISFSIKKVSQKSTKDKLQKETPLTLTKIPOVSLQPAQSG 297  
QY 296 NLTLLAEKTKLHGEVNLVVMRATQKNLTCEVWGPSTPKMLSLKLENKAKVSKRE 355  
DB 298 NLTLLTD--KGLTHGEVNLVVMKVAQNLNTLTCEVWGPSTPKMLTLKQENGEARVSEQ 355  
QY 356 KPVVTLNPDAGMQLSLSDSGQVLESNIKVL 387  
DB 356 KVVQVVAPEGTGLMQLTSEGDVKVMDSRIOVL 387

RESULT 17  
G3MSM  
Ig gamma-3 chain C region, membrane-bound form - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 13-Aug-1986 #sequence\_revision 31-Mar-1991 #text\_change 16-Jul-1999  
A/Accession: A02156; A02155  
R.Wels, J.A.; Word, C.J.; Rimm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blact  
EMBO J. 3, 2041-2046, 1984  
A/Title: Structure analysis of the murine IgG3 constant region gene.  
A/Reference number: A02156; MUID:85027161; PMID:6092053  
A/Accession: A02156  
A/Molecule type: DNA  
A/Residues: 1-398 <WEL>  
A/Cross-references: GB:J00451; NID:G194392; PIDN:AB59655.1; PID:G194433  
R.Komarov, M.; Clayton, L.; Rogers, J.; Robertson, S.; Kettman, J.; Wall, R.  
Nucleic Acids Res. 11, 6775-6785, 1983  
A/Title: The structure of the mouse immunoglobulin in gamma-3 membrane gene segment.  
A/Reference number: A02155; MUID:84041483; PMID:6314258  
A/Accession: A02155  
A/Molecule type: DNA  
A/Residues: 328-332, 'G', 334-341, 'Q', 343-387, 'F', 389-398 <KOM>  
A/Cross-references: GB:K00688  
A/Note: the sequence was determined from the germline gene  
C/Genetics: 97/1; 113/1; 223/1; 328/1; 371/3  
C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap  
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la  
C/Superfamily: immunoglobulin C region; immunoglobulin homology  
C/Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobu  
F.119-83/Domain: immunoglobulin homology <IM1>  
F.136-205/Domain: immunoglobulin homology <IM2>  
F.242-309/Domain: immunoglobulin homology <IM3>  
F.346-362/Domain: transmembrane #status predicted <TM>  
F.363-398/Domain: intracellular #status predicted <INT>  
F.1179,322/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 27.8%; Score 947.5; DB 1; Length 398;  
Best Local Similarity 60.1%; Pred. No. 7.2e-49;  
Matches 172; Conservative 44; Mismatches 49; Indels 21; Gaps 2;

QY 382 SNIKVLPWTSPPVEKSCDKTH-----TCPP---CPAPELLGGPSVVF 420  
DB 64 SLLVVPSPSTPSQTVICNVAHNPASKTELIRIEPRIPKSTPPSSCPGNIIGGSPSVF 123  
QY 421 LFPPKPKDTLMSIRPEVTCVVVDVSHEDPEVKFWMYVDGVNNAKTKKPREBOYNSTYR 480



```

Db      124 IFPPPKALMISLTPKKTCCVVDVSEDDPVHVSFMDNKEVHTAMQPREADYNSTFR 183
      481 VSVSLTVLHDQMLNGKEKCYKSNKALPAPIEKTISKAKGPREQVYTLPPSSDELTKN 540
      184 VVSLALPIHQDMNGKCKCYKSNKALPAPIEKTISKAKGACQVPTTIPPEEQNSKK 243
      541 QVSLTCLVKGFFVPSDIAVEMESNGQPENNYKTTPEVLDSDGSPFLYSKLTVDKSRWQGN 600
      244 KVSILTCLVTLNPFSEALISYEMERNGELFEDYKNTPTILDSGTFLYSKLTVDTSWLGGE 303
      601 VFSGSVMEALHNHYTQKSLSLSPQLDFTCAEAQDGLGLMTT 646
      304 IFTGSVMEALHNHYTQKSLSLSPQLDFTCAEAQDGLGLMTT 349

```

## RESULT 18

```

GIMSK
Ig gamma-1 chain C region, membrane-bound form - mouse
C/Species: Mus musculus (house mouse)
C/Date: 17-Dec-1982 #sequence revision 31-Mar-1991 #text_change 16-Jul-1999
C/Accession: B02159; MUID:80045036; PMID:115593
R/Honjo, T.; Ohta, M.; Yamawaki-Kataoka, Y.; Kataoka, T.; Kawakami, T.; Takahashi, N.;
Cell 18, 559-568, 1979
A/Title: Cloning and complete nucleotide sequence of mouse immunoglobulin gamma1 chain g
A/Reference number: A02159; MUID:80045036; PMID:115593
A/Accession: B02159
A/Molecule type: DNA
A/Residues: 1-393 <HON>
A/Cross-references: GB:J00453
A/Note: The sequence was determined from the germ-line gene
R/Tyler, B.M.; Cowman, A.F.; Geronakakis, S.D.; Adams, J.M.; Bernard, O.
Proc. Natl. Acad. Sci. U.S.A. 79, 2008-2012, 1982
A/Title: mRNA for surface immunoglobulin gamma chains encodes a highly conserved transme
A/Reference number: A02160; MUID:82197626; PMID:6804950
A/Accession: A02160
A/Molecule type: mRNA
A/Residues: 323-393 <TYL>
R/Rogers, J.; Choi, E.; Souza, L.; Carter, C.; Word, C.; Kuehl, M.; Eisenberg, D.; Wall,
Cell 26, 19-27, 1981
A/Title: Gene segments encoding transmembrane carboxyl termini of immunoglobulin gamma c
A/Reference number: A02158; MUID:82115295; PMID:6799207
A/Accession: B02158
A/Molecule type: DNA
A/Residues: 323-366 <ROG>
A/Note: this sequence is the translation of the first exon of the M segment
C/Comment: Cell lines producing IgG contain two mRNA species for Ig gamma chains. The m
ncode membrane-bound chains in that it contains an alternative 3' end, encoded in separa
C/Genetics:
A/Introns: 1/1; 98/1; 111/1; 218/1; 323/1; 366/3
C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob
F/131-200/Domain: immunoglobulin homology <IMM>
F/340-357/Domain: transmembrane #status predicted <TMM>
F/358-393/Domain: intracellular #status predicted <INT>
F/174,278/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match      27.5% Score 940.5; DB 1; Length 393;
Best Local Similarity 63.2% Pred. No. 1.8e-48;
Matches 165; Conservative 46; Mismatches 45; Indels 5; Gaps 2;

      388 PTWSTPEPKSCDTHNCP--CPAPELLGGSVFLFPKPKDTLMISRPEVTCVVVDV 445
      87 PASSTKDKKTLVPRDCCGKPCICTVPEV--SSVFIFPKPKDVLITLTLPKVCVVVDI 143
      446 SHEDPEVKFNMYVDGVEVHNAKTKPREQYNSTYRVSVLTTLVLDQMLNGKEKCYKSNK 505
      144 SKDDPEVQFSWFDVDEVHTAQTPREQFNSTFRSVSELPIMQDMLNGKEKFCRNSA 203
      506 ALPAPIEKTISKAKGPREQVYTLPPSRDELTKNQVSLTCLVKGFFPSDIAVEMESNGQ 565

```

```

Db      204 AFPAPIEKTISKGRKPAQVYTTIPPEKQMAKDKVSLTCMTDFPEPDIYEMQNGQ 263
      566 PENNYKTTPTVLDSDGSPFLYSKLTVDKSRWQGNVPSCSVMEALHNHYTQKSLSPG 625
      264 PAENYKTKQPIIMNTNGSYFYSKLTNQGKSNMEAGNTTCVSLHGLHNHTKSLSPG 323
      626 LQDFTCAEAQDGLGLMTT 646
      324 LQDFTCAEAQDGLGLMTT 344

```

## RESULT 19

```

GHRB
Ig gamma chain C region - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 24-Apr-1984 #sequence revision 15-Nov-1984 #text_change 16-Jul-1999
C/Accession: A91749; A90290; A93928; A90245; A94416; A02161
R/Bernstein, K.E.; Alexander, C.B.; Mage, R.G.
Immunogenetics 18, 387-397, 1983
A/Title: Nucleotide sequence of a rabbit IgG heavy chain from the recombinant F-1 haplocl
A/Reference number: A91749; MUID:84030930; PMID:6313520
A/Accession: A91749
A/Molecule type: mRNA
A/Residues: 1-323 <BER>
A/Note: This sequence has the d12 allotypic marker, 104-Thr, and the e14 marker, 185-Thr
R/Pratt, D.M.; Mole, L.E.
Biochem. J. 151, 337-349, 1975
A/Title: Sequence studies on the constant region of the Fd sections of rabbit immunoglob
A/Reference number: A90290; MUID:76135465; PMID:1243651
A/Accession: A90290
A/Molecule type: Protein
A/Residues: 1-47, 'E', '49-71', 'PV', '72-128 <PRA>
R/Martens, C.L.; Moore, K.W.; Steinmetz, M.; Hood, L.; Knight, K.L.
Proc. Natl. Acad. Sci. U.S.A. 79, 6018-6022, 1982
A/Title: Heavy chain genes of rabbit IgG: isolation of a cDNA encoding gamma heavy chain
A/Reference number: A93928; MUID:8329917; PMID:6193512
A/Accession: A93928
A/Molecule type: mRNA
A/Residues: 88-103, 'W', '105-143', 'E', '145-184', 'A', '186', 'E', '188-266 <MAR>
A/Cross-references: GB:M16426; NID:g165111; PIDN:AA31289.1; PID:g165112
A/Note: this sequence has the d11 allotypic marker, 104-Met, and the e15 allotypic marker
R/Fruehner, R.G.; Jackson, S.A.; Mole, L.E.; Porter, R.R.
Biochem. J. 116, 249-259, 1970
A/Title: Sequence studies of the Fd section of the heavy chain of rabbit immunoglobulin (
A/Reference number: A90245; MUID:70110015; PMID:5461106
A/Accession: A90245
A/Molecule type: Protein
A/Residues: 132-143, 'E', '145-161 <FRU>
R/Hill, R.L.; Lebovitz, H.E.; Fellows Jr., R.E.; Delaney, R.
in Gamma Globulins, Nobel Symp. 3, Killander, J., ed., pp.109-127, Almqvist and Wiksell,
A/Reference number: A94416
A/Accession: A94416
A/Molecule type: Protein
A/Residues: 159-151;155-172, 'D', '174-184', 'A', '186', 'E', '188-200', 'D', '202-217', 'E', '219-232', 'Q', '
C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
F/20-82/Domain: immunoglobulin homology <IM1>
F/130-199/Domain: immunoglobulin homology <IM2>
F/236-303/Domain: immunoglobulin homology <IM3>
F/173/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match      27.0% Score 922; DB 1; Length 323;
Best Local Similarity 69.3% Pred. No. 1.8e-47;
Matches 169; Conservative 30; Mismatches 37; Indels 8; Gaps 2;

      382 SNIKVLPWTWSTPEPKSCDTHNCP--CPAPELLGGSVFLFPKPKDTLMISRPEVTCV 441
      87 TTKTKVDKT-----VAPSTCKR-----PTCPPPELLGGSVFLFPKPKDTLMISRPEVTCV 138
      442 VVDVSHEDPEVKFNMYVDGVEVHNAKTKPREQYNSTYRVSVLTTLVLDQMLNGKEKCK 501

```



```

Db      139 VVDVSDQDDPEVQFTWYINNEQVRRTPARPELRQGFNFSTIRRVSTLPITPHQDLRLKGEFFCK 198
Oy      502 VSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAWE 561
Db      199 VHNKALPAPIEKTISKAKGQPLEPKVTGMGPREELSSRSVSLTCMINGFYPDSIWE 258
Oy      562 SNGQEPNNYKTTPTPLVDSGSPFLYXSKLTVDKSRHQGNVSSCSVMHEALHNHYTQKSLS 621
Db      259 KNGKAEKDYKTTTPAVLDSGSSYFLYNNKLSVPTSEWRGDVFTCSVMHEALHNHYTQKSIS 318
Oy      622 LSPG 625
Db      319 RSPG 322

RESULT 20
G2MSM
19 gamma-2a chain C region, membrane-bound form - mouse
C|Species: Mus musculus (house mouse)
C|Dates: 19-Feb-1994 #sequence revision 31-Mar-1991 #text_change 16-Jul-1999
C|Accession: A02154; B32657; I57809
R|Yamawaki-Kataoka, Y.; Nakai, S.; Miyata, T.; Honjo, T.
P|Proc. Natl. Acad. Sci. U.S.A. 79, 2623-2627, 1982
A|Title: Nucleotide sequences of gene segments encoding membrane domains of immunoglobulin
A|Reference number: A02154; MUID:82222190; PMID:6283537
A|Accession: A02154
A|Molecule type: DNA
A|Residues: 329-399 <YAM>
A|Cross-references: GB:J00471
A|Notes: the sequence was determined from the germ-line gene
R|Yamawaki-Kataoka, Y.; Miyata, T.; Honjo, T.
Nucleic Acids Res. 9, 1365-1381, 1981
A|Title: The complete nucleotide sequence of mouse immunoglobulin gamma 2 A gene and evoc
A|Reference number: A32657; MUID:81198976; PMID:6262729
A|Accession: B32657
A|Molecule type: DNA
A|Residues: 1-329, 'K' <YA2>
R|Hall, B.; Mlicarek, C
MOL. Immunol. 26, 819-826, 1989
A|Title: Sequence and polyadenylation site determination of the murine immunoglobulin gamma
A|Reference number: I57809; MUID:90097953; PMID:2513486
A|Accession: I57809
A|Status: preliminary; translated from GB/EMBL/DBJ
A|Molecule type: DNA
A|Residues: 373-399 <RES>
A|Cross-references: GB:M5032; NID:g194478; PIDN:AAA3791.1; PID:g387217
C|Comment: The sequence of residues 1-328 was assumed to be identical with the correspond
C|Comment: Cell lines producing I9G contain two mRNA species for Ig gamma chains. The ma
hat it contains an alternative 3' end, encoded in separate exons, that is homologous with
C|Genetics:
A|Intons: 1/1, 98/1, 114/1, 224/1, 329/1, 372/1
C|Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap
C|Chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
C|Superfamily: immunoglobulin C region; immunoglobulin homology
C|Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobul
F|137-206/Domain: immunoglobulin homology <IMW>
F|346-353/Domain: transmembrane #status predicted <TM>
F|364-399/Domain: intracellular #status predicted <INT>
F|180/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match      26.8%; Score 915.5; DB 1; Length 399;
Best Local Similarity 49.6%; Pred. No. 5.7e-47;
Matches 183; Conservative 48; Mismatches 79; Indels 59; Gaps 7;

Oy      285 PQLPQVAGSNLTALAEAKTGKHQEV-----NLVWRATQLOKNTLCEYWGFTSPKLM 339
Db      34  PEPYTLTNWSSGLSSGVTFPAVLQSDLYLTSSSVTVTSSWTPQSDITCNVAHPAS----- 89
Oy      340 LSLKLENKAVSYSRKREKVVWVLNPEACGMQGLLSDSGVLLSEIKYVLPWTSTVEPKSC 399
Db      90  -----STVWDKIEP-----RG-----PT---IKP--- 106

```

Query Match 26.7%; Score 911.5; DB 2; Length 328;  
Best Local Similarity 63.1%; Pred. No. 7.6e-47;  
Matches 173; Conservative 34; Mismatches 46; Indels 21; Gaps 4;

Query 371 LLSDSGQVLLSEINIKVLPWSTPVEPKSCDKTH-----TCPPCPAPDEL 413  
Db 56 VLPSPGLVSLSMWTV-PASSLSKSYTCNVNHPATTKYDKRGVGTIKTKPCPICPACE- 113  
414 LGSPSVFLPPPKPKDTLMSRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPRRE 473  
Db 114 SPGPSVFIFPFPKDTLMSRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPRRE 173  
474 QYNSTRYRVSVLTVLHODMLNGEKYCKYSNKALPAPIEKTISAKAKQPREPQYTTLPSS 533  
Db 174 QNSTYRVSVSVLPIDHODMLNGEKYCKYSNKALPAPIEKTISAKAKQPREPQYTTLPSS 233  
534 RELTKNQVSLTCLVGVGFPYSPIADVEMESNGQ--PENNYTKTPPVLSDSGSFPLYSKLTV 591  
Db 234 AEELRSKSYITCLVIGFIPPDIDVEMQNRGQEPESNRYITTPPQDQVDGTYFLYKPSV 293  
592 DKSRWQGNVFSQVMEALHNNHYTKSLSLSPG 625  
Db 294 DKASWQGGIFQCAVVMHEALHNNHYTKSLSLSPG 327

RESULT 21  
147160  
Ig gamma 2b chain constant region - pig (fragment)  
C1Species: Sus scrofa domestica (domestic pig)  
C1Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 21-Jan-2000  
C1Accession: 147160  
R1Kacskovics, I.; Sun, J.; Butler, J.E.  
A1Title: Five putative subclasses of swine Igg identified from the cdna sequences of a s  
A1Reference number: 147158; MID:95015845; PMID:7930579  
A1Accession: 147160  
A1Status: preliminary; translated from GB/EMBL/DBJ  
A1Molecule type: mRNA  
A1Residues: 1-328 <KAC>  
A1Cross-references: EMBL:U03780; NID:g433125; PIDN:AA52218.1; RID:g433126  
C1Genetics:  
C1Gene: IgG2b  
C1Superfamily: immunoglobulin C region; immunoglobulin homology  
F133-202/Domain: immunoglobulin homology <IMM>

Query Match 26.7%; Score 911.5; DB 2; Length 328;  
Best Local Similarity 63.1%; Pred. No. 7.6e-47;  
Matches 173; Conservative 34; Mismatches 46; Indels 21; Gaps 4;

Query 371 LLSDSGQVLLSEINIKVLPWSTPVEPKSCDKTH-----TCPPCPAPDEL 413  
Db 56 VLPSPGLVSLSMWTV-PASSLSKSYTCNVNHPATTKYDKRGVGTIKTKPCPICPACE- 113  
414 LGSPSVFLPPPKPKDTLMSRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPRRE 473  
Db 114 SPGPSVFIFPFPKDTLMSRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPRRE 173  
474 QYNSTRYRVSVLTVLHODMLNGEKYCKYSNKALPAPIEKTISAKAKQPREPQYTTLPSS 533  
Db 174 QNSTYRVSVSVLPIDHODMLNGEKYCKYSNKALPAPIEKTISAKAKQPREPQYTTLPSS 233  
534 RELTKNQVSLTCLVGVGFPYSPIADVEMESNGQ--PENNYTKTPPVLSDSGSFPLYSKLTV 591  
Db 234 AEELRSKSYITCLVIGFIPPDIDVEMQNRGQEPESNRYITTPPQDQVDGTYFLYKPSV 293  
592 DKSRWQGNVFSQVMEALHNNHYTKSLSLSPG 625  
Db 294 DKASWQGGIFQCAVVMHEALHNNHYTKSLSLSPG 327

RESULT 21  
147160  
Ig gamma 2b chain constant region - pig (fragment)  
C1Species: Sus scrofa domestica (domestic pig)  
C1Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 21-Jan-2000  
C1Accession: 147160  
R1Kacskovics, I.; Sun, J.; Butler, J.E.  
A1Title: Five putative subclasses of swine Igg identified from the cdna sequences of a s  
A1Reference number: 147158; MID:95015845; PMID:7930579  
A1Accession: 147160  
A1Status: preliminary; translated from GB/EMBL/DBJ  
A1Molecule type: mRNA  
A1Residues: 1-328 <KAC>  
A1Cross-references: EMBL:U03780; NID:g433125; PIDN:AA52218.1; RID:g433126  
C1Genetics:  
C1Gene: IgG2b  
C1Superfamily: immunoglobulin C region; immunoglobulin homology  
F133-202/Domain: immunoglobulin homology <IMM>











## RESULT 27

S22080

Ig heavy chain precursor (B/MT.4A.17.H5.A5) - bovine

N/Alternate names: Ig gamma-1 chain C region (clone 8.10)

C/Species: Bos primigenius taurus (cattle)

C/Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text\_change 23-Jul-1999

C/Accession: S22080; S06610; A31303

R/Sanders, P.G.

submitted to the EMBL Data Library, November 1991

A/Reference number: S22080

A/Accession: S22080

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-470 &lt;S&amp;N&gt;

A/Cross-references: EMBL:X62916; NID:9439; PID:CAA44699.1; PID:9440

R/Symons, D.B.A.; Clarkson, C.A.; Beale, D.

Mol. Immunol. 26, 841-850, 1989

A/Title: Structure of bovine immunoglobulin constant region heavy chain gamma 1 and gamma 2

A/Reference number: S06610; MUID:90097956; PMID:2513487

A/Accession: S06610

A/Molecule type: DNA

A/Residues: 142-470 &lt;SYM&gt;

A/Cross-references: EMBL:X16701

A/Note: the sequence was determined from the germline gene

C/Genetics:

A/Introns: 98/1; 111/1; 221/1

C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: glycoprotein; heterotrimer; immunoglobulin; membrane protein

F/161-225/Domain: immunoglobulin homology &lt;IMM&gt;

F/318/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 25.9%; Score 884.5; DB 2; Length 470;

Best Local Similarity 35.6%; Pred. No. 4.8e-45;

Matches 227; Conservative 58; Mismatches 164; Indels 189; Gaps 16;

Db 6 PFRHLVLVQ-----LALLPAATQGNKVVLAGKGGVTLCTAS--QKSIQFMKSNQ 58

Db 3 PLMTLLFLVLSAPIGVLSVQVLRSGPSLV--KPSQTLSTLTCTVSGFSLSTALTWVROAP 60

Db 59 IKILGNOSFLTKPKSKLNDRAISRSLMDGN--PFLIKIKLIEDSDTYIC----- 109

Db 61 GKALEWVGITSGGTYVNPALKSRSLSTKENSQSVLSVSTPBDTATVYCARSTYG 120

Db 110 EVEDQKEVQLVGLTANSTHLLQGSLLTLTSPSGSSPQCSPRKATIGGKTL 169

Db 121 EVGG-----GAIADA-----WGGGLVTVSSASTTAPKYPLS---SCCGKSS 161

Db 170 SVSGLQLDQSGTWTCTVLQNGKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTV 229

Db 162 STVTL-----GCLVSYNPEPTV 180

Db 230 EKLTSGLMLWQABRASSKSMITFDLKNKEVSVKRVYTDPKLQWKKLPLHLYLPOALP 289

Db 181 -----TMSGALK----- 188

Db 290 QYASSGNLTALAEKTLGHLQBNLVYWRATQLOKNTLCEVWGPSPKMLSLKNEKA 349

Db 189 ---SGVTFPAVLIQSSGLYSLSNMVTPGSTQGTFCNVAHPAS-----ST 232

Db 350 KVSREKPVVWLVNPDAGWQCLLSDSGVLLSNIKVLPTWSTVEPEKSCDKHTCPPCP 409

Db 233 KVDADVPT-----CKSPCD---C-CP 251

Db 410 APELLGGSPVLPFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTK 469

Db 252 PPELPGGSPVLPFPKPKDTLTISGTEPVCVVVDVGDHDEPKVFSWVDVDEVTATTK 311

Db 470 PREQVNSTYRVVSVLVTLHODMLNGKSKYKCNKALPAIEKTIISKAKQPREPQYTT 529

Db 312 PREQVNSTYRVVSVLVTLHODMLNGKSKYKCNKALPAIEKTIISKAKQPREPQYTT 371

## RESULT 28

I47161

Ig gamma 3 chain constant region - pig (fragment)

C/Species: Sus scrofa domestica (domestic pig)

C/Date: 21-Feb-1997 #sequence revision 21-Feb-1997 #text\_change 21-Jan-2000

C/Accession: I47161

R/Kacskovics, I.; Sun, J.; Butler, J.B.

J. Immunol. 153, 3565-3573, 1994

A/Title: Five putative subclones of swine Igg identified from the cDNA sequences of a s

A/Reference number: I47158; MUID:95015845; PMID:7930579

A/Accession: I47161

A/Status: preliminary; translated from GB/EMBL/DDBJ

A/Molecule type: mRNA

A/Residues: 1-328 &lt;KAC&gt;

A/Cross-references: EMBL:U03781; NID:9433127; PID:AAA52219.1; PID:9433128

C/Genetics:

A/Introns: 193

C/Superfamily: immunoglobulin C region; immunoglobulin homology

F/133-202/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 25.9%; Score 883.5; DB 2; Length 328;

Best Local Similarity 62.0%; Pred. No. 3.4e-45;

Matches 170; Conservative 32; Mismatches 51; Indels 21; Gaps 4;

Db 371 LLSDSGVLLSNIKVLPTWSTVEPEKSCDKHT-----TCPECPAPEL 413

Db 56 VLPSPGLYSLSNMVTP--PASLSKSKYTCNVNHPATTYKDKRVGTYKPPCPICPCE- 113

Db 414 LGGSPVLPFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKPRE 473

Db 114 VAGPSVLPFPKPKDTLMISQTEPEVTCVVVDVSKENAEVPSKYVDGVEVHTATRKKE 173

Db 474 QYNSTYRVVSVLVTLHODMLNGKSKYKCNKALPAIEKTIISKAKQPREPQYTTLPS 533

Db 174 QPNSTYRVVSVLPVQHDMVKGKFKCKVANNVDPARTITISKALQGSREPPQYTLPP 233

Db 534 RDELTKQVSLTCLVKGFPYSDIAVEMESNGO--PENNYKTTTPVLDSDGSPFLYSKLT 591

Db 234 AEELSRKSVTVTLVIGFYPPDIHVEKWSNGQPEPEGNRYRTTPQDDVDGTFPLYSKLAV 293

Db 592 DKSRWQGNVFGSCVMHEALHNHYTKSLSPG 625

Db 294 DKARMDHGETFECAVMHEALHNHYTKSKISKTG 327

## RESULT 29

S31459

Ig gamma-1 chain - sheep (fragment)

C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C/Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text\_change 16-Jul-1999

C/Accession: S31459

R/Patri, S.; Nau, F.

submitted to the EMBL Data Library, December 1992

A/Reference number: S31459

A/Accession: S31459

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-472 &lt;PAT&gt;

A/Cross-references: EMBL:X69797

C/Superfamily: immunoglobulin C region; immunoglobulin homology

F/277-346/Domain: immunoglobulin homology &lt;IMM&gt;



[illegible]

Db 152 EQGFNSTFRVVSALPLQHDWDTGKGFCKVHNHEALPAPVIRITISRKKGARREQVYTLA 211

QY 532 PSRDLITKQVSLTCLVKGFPSPDIIVEMESNGQP--ENNYKTPPVLDSGSEFFLYSKL 589

Db 212 PQGEISLSTSLTCLVGTGFPYDPIYAVEMQKNGPSESDKXGTTTSGLDADGSEFLYSRL 271

QY 550 TYDKSRWQGNVFSQGVHAEALHNHYTKSLSPG 625

Db 272 RVDKNSWQGDYTYACVHAEALHNHYTKSLSPG 307

RESULT 31

G3MSC

Ig gamma-3 chain C region, secreted form - mouse

C:Species: Mus musculus (house mouse)

C:Date: 17-Mar-1987 #sequence\_revision 31-Mar-1991 #text\_change 16-Jul-1999

C:Accession: B02156

R:Wels, J.A.; Word, C.J.; Rimm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blatt

EMBO J. 3, 2041-2046, 1984

A:Title: Structure analysis of the murine IgG3 constant region gene.

A:Reference number: A02156; MUID:85027161; PMID:6092053

A:Accession: B02156

A:Molecule type: DNA

A:Residues: 1-329 <HEL>

A:Cross-references: GB:J00451

A:Note: the sequence was determined from the germline gene

C:Genetics:

A:introns: 97/1; 113/1; 223/1

C:Complex: An immunoglobulin heterodimer subunit consists of two identical light (kap

hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: alternative splicing; duplication; glycoprotein; heterodimer; immunoglob

F:19-83/Domain: immunoglobulin homology <IM1>

F:97-112/Region: hinge

F:136-205/Domain: immunoglobulin homology <IM2>

F:242-309/Domain: immunoglobulin homology <IM3>

F:179,322/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 24.9%; Score 849.5; DB 1; Length 329;

Best Local Similarity 58.1%; Pred. No. 3, 5e-43;

Matches 154; Mismatch 42; Mismatches 48; Indels 21; Gaps 2;

QY 362 SNIKVLPWSPVPEPKSCDKTH-----TCPD--CPABELLGPSVF 420

Db 64 SSLVTVPSSTWPSQVYICVAVHPASKTELKRIEPRIPKSTSPGSSCPGNITLGPSVF 123

QY 421 LPPPKDGLTMSRIPEVTCVVYDVSHDEPEKFMNYVDGVENATKTPREEQYNSTYR 480

Db 124 IFPPPKALMTSLTPKVTCCVVDVSDPDVHVSFWDKKEVNTAQTQPREAQNSTFR 183

QY 481 VSVTVLTHQDLNGKEYKCVSNKALPAPIETKISAKGQPREPOVYTPPSPDELTKN 540

Db 184 VSAALPIQHDMMRGKEFKCKVNNKALPAPIETKISAKGQPREPOVYTPPSPDELTKN 243

QY 541 QVSLTCLVKGFPSPDIIVEMESNGQPENNYKTPPVLDSGSEFFLYSKLTYDKSRWQGN 600

Db 244 KVSLLCLVNTFNSSAISVEMERGELGQYKNTPTPLDSGTYFLYSKLTVDTSWLGGE 303

QY 601 VFGSGVHAEALHNHYTKSLSPG 625

Db 304 IFTCGVHAEALHNHYTKSLSPG 328

RESULT 32

PS0018

Ig gamma-2b chain C region - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 16-Jul-1999

C:Accession: PS0018; B25941

R:Bruggemann, M.

Gene 74, 473-482, 1988

A:Title: Evolution of the rat immunoglobulin gamma heavy-chain gene family.

A:Reference number: PS0017; MUID:89232738; PMID:3149946



A:Accession: P50018  
A:Molecule type: DNA  
A:Residues: 1-333 <BRU>  
R:Brueggemann, M.; Free, J.; Diamond, A.; Howard, J.; Cobbold, S.; Waldmann, H.  
Proc. Natl. Acad. Sci. U.S.A. 83, 6075-6079, 1986  
A>Title: Immunoglobulin heavy chain locus of the rat: striking homology to mouse antibody  
A:Reference number: A5941; MUID:86287397; PMID:3016742  
A:Accession: B25941  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 227-333 <BR2>  
C:Genetics:  
A:Introns: 96/1; 117/1; 227/1  
C:Superfamily: Immunoglobulin C region; immunoglobulin homology  
C:Keywords: Immunoglobulin  
F:20-82/Domain: immunoglobulin homology <IM>

Query Match 24.7%; Score 844; DB 2; Length 333;  
Best Local Similarity 61.7%; Pred. No. 7.6e-43;  
Matches 153; Conservative 34; Mismatches 51; Indels 10; Gaps 1;

QY PTWSTPVEPK-----SCDKTHTCPCPAPBLGGPSVLPFPKPKDTLMISRPE 437  
DB PASSTKVKKYBRNNGIGHKCPTCTCHKCPVELLGGPSVFIFPPKPDILISQVAK 144  
QY 438 VTCVVVDVSHEDPEVKFNNYVGVENNAKTKPREEOYNSTYRVSVLTVLHODMNGKE 497  
DB 145 VTCVVVDVSEEDPVQFSFVNNEVHTAQTQPREOYNSTFRVVSALPIQHODMNGKE 204  
QY 498 YKCKVSKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKNOVSLTCLVKGYPSPDIA 557  
DB 205 FKCKNNVALPSPLEKITSKRGGLVKKPQVYVVGPTQLTEQVLSLCLTSGFLPNDIG 264  
QY 558 VEWESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSKNOGQNVSCGMHEALHNHYTQ 617  
DB 265 VEWISNGIEKQYKTEPRVMDSDGSFMYSKLNVKRSKMSDRAPVCSVVEGLHNHVE 324  
QY 618 KSLSLSPG 625  
DB 325 KSLSRPFG 332

RESULT 33  
PC4436  
monoclonal antibody 13-1 heavy chain - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 04-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 21-Jan-2000  
C:Accession: PC4436  
R:Akashi, S.; Kato, K.; Torizawa, T.; Dolmae, N.; Yamaguchi, H.; Kamachi, M.; Harada, A.  
Biochem. Biophys. Res. Commun. 240, 566-572, 1997  
A>Title: Structural characterization of mouse monoclonal antibody 13-1 against a porphyrin  
A:Reference number: J5810; MUID:98063277; PMID:9398605  
A:Accession: PC4436  
A:Molecule type: protein  
A:Residues: 1-444 <AKA>  
C:Comment: This catalytic antibody has peroxidase oxidase activity. It is directed against  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
F:251-330/Domain: immunoglobulin homology <IM>  
F:122/Disulfide bonds: interchain (to 98) #status predicted  
F:99/Disulfide bonds: interchain (to 109) #status predicted

Query Match 24.6%; Score 839.5; DB 2; Length 444;  
Best Local Similarity 60.8%; Pred. No. 2e-42;  
Matches 146; Conservative 44; Mismatches 45; Indels 5; Gaps 2;

QY 388 PTWSTPVEPKSCDKTHTCPP--CPAPBLGGPSVLPFPKPKDTLMISRPEVTCVVVDV 445  
DB 207 PASSTKVKKLVPRDCCGKPCICTVPEV---SSVFIFPPKPKDVLTLITLTKVTCVVVDI 263  
QY 446 SHEDEVEKNNYVGVENNAKTKPREEOYNSTYRVSVLTVLHODMNGKEKYGKYSNK 505  
DB 264 SKDPEVQFSFVNNEVHTAQTQPREOYNSTFRVVSALPIQHODMNGKEFKCRVNSA 323

QY 506 ALPAPIEKTISKAKGPREPOVYTLPPSRDELTKNOVSLTCLVKGYPSPDIAMWESNGQ 565  
DB 324 APAPIEKTISKTKGRKAPQVYTLIPPEKQAMKDKVSLTCMTIDFPEDITVEMQNNQ 383  
QY 566 PENNYKTPPVLDSDGSFPLYSKLTVDKSKNOGQNVSCGMHEALHNHYTQSLSPG 625  
DB 384 PAENYKTPQIPMDTDSGFYVYSKLVNOKSWMEAGNFTCSVLHGLHNHHTKSLSPG 443

RESULT 34  
S37483  
Ig gamma-2a chain - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999  
A:Accession: S37483  
R:Duncanson, F.F.D.  
submitted to the EMBL Data Library, February 1993  
A:Reference number: S37483  
A:Accession: S37483  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-469 <DUC>  
A:Cross-references: EMBL:X70423; NID:g406252; PID:CA49868.1; PID:g406253  
C:Superfamily: Immunoglobulin C region; immunoglobulin homology  
C:Keywords: Immunoglobulin  
F:276-345/Domain: immunoglobulin homology <IM>

Query Match 24.5%; Score 836.5; DB 2; Length 469;  
Best Local Similarity 35.4%; Pred. No. 3.3e-42;  
Matches 216; Conservative 66; Mismatches 142; Indels 187; Gaps 21;

QY 30 LGKKGDVVELTCTASQKSIQFPMKNSNQIKLGNQ-----SFLTGPSKLNDRADSR 83  
DB 30 LVPRGASVKSICASGVTFTDY--INWYKQPGQGLKMIWYIPASGATKYNENKKG 86  
QY 84 RSL---WDGNNFLLIKNLKIEDSDTYICEVDQKEEVQLVFGLTANSPTHLIQ--GQS 138  
DB 87 ATLTVDTSSSTAYMQLSLTSEDTAVYFC-----ARANGATAT---LLDYWGQG 132  
QY 139 LTLTLESPPSGSPQCRSGKNIQGGKTLVSQLELDQDSGTWTCVTLQNKVVEKID 198  
DB 133 TLTIVSSAKTTAPSVPLAP---VCGDTT----- 158  
QY 199 IVVLAPEKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMQAERASSSKSWITPDLKN 258  
DB 159 -----GSSVTL---GCLVKGYPF---EYVT---LTWNSGSLSSG----- 188  
QY 259 KEYSVKRVITODPKLQNGKKLPLHLTLPLQLPQ--YAGSGNLTALAEKTKLHOENVLV 316  
DB 189 -----VH--TEPAVLQSDLYTLSSSVT-----V 209  
QY 317 MRATQLOKNTLCVWGPTSPKMLSLKLEKKAIVSRKRPVWVYLNPEAGMCCLSDSG 376  
DB 210 TSSTPSQSITTCVANHAS-----STKVDKTEP-----RG 240  
QY 377 QVLLSNIKVLPWSTPVEPKSCDKTHTCPP--CPAPBLGGPSVLPFPKPKDTLMISR 434  
DB 241 -----PFI---IKP-----CPKCPAPNLLGGPSVFIFPPKIDVLMISL 277  
QY 435 TPVYTCVVVDVSHEDPEVKFNNYVGVENNAKTKPREEOYNSTYRVSVLTVLHODMNL 494  
DB 278 SPVITCVVDVSDDDPDVQISWFFVNNEVHTAQTQPREOYNSTLRVVSALPIQHODMNS 337  
QY 495 GKXEKCVSKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKNOVSLTCLVKGYPSP 554  
DB 338 GKPEFKCVNNKDPAPBERTISKPKGSVRAPOVYVLPPEEEMTKQVTLTCMTIDMPE 397  
QY 555 DIAMWESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSKNOGQNVSCGMHEALHNH 614  
DB 398 DIIVEMTNNGKTLANTKNTERTPVLDSDGSYFMYSKLVKVEKKWVERNSYSCSVVEGLHNH 457  
QY 615 YTKSLSLSPG 625



Db 458 HTTKSFSPRTPG 468

## RESULT 35

PS0017

Ig gamma-1 chain C region - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 16-Jul-1999

C:Accession: PS0017; C25941

R:Bruggemann, M.

Gene 74, 473-482, 1988

A:Title: Evolution of the rat immunoglobulin gamma heavy-chain gene family.

A:Reference number: PS0017; MUID:89232738; PMID:3149946

A:Accession: PS0017

A:Molecule type: DNA

A:Residues: 1-326 &lt;BRU&gt;

A:Cross-references: GB:V00775; NID:951652; PIDN:CAA24153.1; PID:G51653

R:Bruggemann, M.; Free, J.; Diamond, A.; Howard, J.; Cobbold, S.; Waldmann, H.

Proc. Natl. Acad. Sci. U.S.A. 83, 6075-6079, 1986

A:Title: Immunoglobulin heavy chain locus of the rat: striking homology to mouse antibody

A:Reference number: A25941; MUID:86287397; PMID:3016742

A:Accession: C25941

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 220-326 &lt;BR2&gt;

C:Genetics:

A:introns: 98/1; 113/1; 220/1

C:Superfamily: Immunoglobulin C region; immunoglobulin homology

C:Keywords: Immunoglobulin

F:20-84/Domain: immunoglobulin homology &lt;IMN&gt;

Query Match 24.3%; Score 828; DB 2; Length 326;

Best Local Similarity 55.1%; Pred. No. 6.6e-42;

Matches 150; Conservative 48; Mismatches 48; Indels 26; Gaps 4;

QY 375 SQGVLLSNIKVLPTWSTVPEKSCDKTH-----TCPPCAPELLGG 416

Db 59 SGLYLTSSVTV-PSSTWPSQVTCNVAPASSTKVDKXIVRNCGGDKPC---ICRG 113

QY 417 ---PSVFLPPPKKDTLMISRTPEVTCVVVDVSHPEPEKFNKYVDGVENNAKTPREE 473

Db 114 SEVSVFIFPPPKDVLITTLTPKVCVVVDISQDDPEVHFSFVVDVEVHTAQTREPRB 173

QY 474 QVNSTRVAVSVLTVLHODMNGKEYCKVSNKALPAPIEKTISKAKGQPREPVYTLPPS 533

Db 174 QFVSTFRSVSELPILHODMNGRTFKCTSAAPSPRIEKTISKPEGRQVPPVYVMSPT 233

QY 534 RDELTKQVSLTCLVKGFPYPSDIAVWESNGQPENNYKTPPYLSDSGSFYLSKLTVDK 593

Db 234 KEEMTQNEVSITCMVAGFPPRDIYVEQMONGQPEHYKTPPTMDIDSGSYFLSKLVNKK 293

QY 594 SRWQGNVSCSVMEALNNHYTKSLSPG 625

Db 294 EKWQGNFTCTCSVLHGLHNNHTEKLSHSPG 325

## RESULT 36

GIMS

Ig gamma-1 chain C region, secreted form - mouse

C:Species: Mus musculus (house mouse)

C:Date: 31-Mar-1980 #sequence\_revision 24-Sep-1981 #text\_change 16-Jul-1999

C:Accession: A02159; A26234; A26236

R:Honjo, T.; Obara, M.; Yamawaki-Kataoka, Y.; Kataoka, T.; Kawakami, T.; Takahashi, N.;

Cell 18, 559-568, 1979

A:Title: Cloning and complete nucleotide sequence of mouse immunoglobulin gamma1 chain g

A:Reference number: A02159; MUID:80045036; PMID:115593

A:Accession: A02159

A:Molecule type: DNA

A:Residues: 1-324 &lt;HON&gt;

A:Cross-references: GB:000453

A:Note: the sequence was determined from the germline gene

A:Note: Lys-324 is removed posttranslationally

R:Obara, M.; Yamawaki-Kataoka, Y.; Takahashi, N.; Kataoka, T.; Shimizu, A.; Mano, Y.; Se

Gene 9, 87-97, 1980

A:Title: Immunoglobulin gamma-1 heavy chain gene: structural gene sequences cloned in a

A:Reference number: A26234; MUID:80202559; PMID:6769752

A:Contents: MOPC 31C

A:Accession: A26234

A:Molecule type: mRNA

A:Residues: 76-324 &lt;OBA&gt;

A:Cross-references: GB:V00775; NID:951652; PIDN:CAA24153.1; PID:G51653

R:Rogers, J.; Clarke, P.; Salsner, W

Nucleic Acids Res. 6, 3305-3321, 1979

A:Title: Sequence analysis of cloned cDNA encoding part of an immunoglobulin heavy chain

A:Reference number: A26236; MUID:80012837; PMID:113776

A:Accession: A26236

A:Contents: MOPC 21

A:Molecule type: mRNA

A:Residues: 170-275 'D', 277 'D', 279-322 &lt;ROG&gt;

A:Cross-references: GB:V00795; NID:951830; PIDN:CAA24176.1; PID:G780265

R:Adelugbo, K.

J. Biol. Chem. 253, 6068-6075, 1978

A:Title: Evolution of immunoglobulin subclasses. Primary structure of a murine myeloma g

A:Reference number: A26237; MUID:78242288; PMID:98524

A:Contents: MOPC 21

A:Note: this is the final paper in a series reporting the protein sequence, the disulfid

A:Note: there are a number of differences from the sequence shown

C:Genetics:

A:introns: 1/1; 98/1; 113/1; 218/1

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap

hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la

C:Superfamily: Immunoglobulin C region; immunoglobulin homology

C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobu

F:20-84/Domain: immunoglobulin homology &lt;IM1&gt;

F:98-110/Region: hinge

F:131-200/Domain: immunoglobulin homology &lt;IM2&gt;

F:237-304/Domain: immunoglobulin homology &lt;IM3&gt;

F:27-82, 138-198, 244-302/Disulfide bonds: #status experimental

F:102/Disulfide bonds: interchain (to light chain) #status experimental

F:104,107,109/Disulfide bonds: interchain (to heavy chain) #status experimental

F:174/Binding site: carbohydrate (asn) (covalent) #status experimental

Query Match 24.2%; Score 826.5; DB 1; Length 324;

Best Local Similarity 60.0%; Pred. No. 8e-42;

Matches 144; Conservative 46; Mismatches 45; Indels 5; Gaps 2;

QY 388 PWSVPEPKSDKNTKRP--CPARELLGSGSVFLFPKPKDTLMISRTPEVTCVVVDV 445

Db 87 PWSVPEPKSDKNTKRP--CPARELLGSGSVFLFPKPKDTLMISRTPEVTCVVVDV 143

QY 446 SHEDEVEKFNMYVDDEVENNAKTPREQVNSTVYVSVLTCLVKGFPYSDIAVWESNGQ 505

Db 144 SKDDEVEKFNMYVDDEVENNAKTPREQVNSTVYVSVLTCLVKGFPYSDIAVWESNGQ 203

QY 506 ALPAPIEKTISKAKGQPREPVYTLPPSRDELTKQVSLTCLVKGFPYSDIAVWESNGQ 565

Db 204 AFPAPIEKTISKAKGQPREPVYTLPPSRDELTKQVSLTCLVKGFPYSDIAVWESNGQ 263

QY 566 PENNYKTPPYLSDSGSFYLSKLTVDKSRWQGNVSCSVMEALNNHYTKSLSPG 625

Db 264 PAENYKTPPYLSDSGSFYLSKLTVDKSRWQGNVSCSVMEALNNHYTKSLSPG 323

## RESULT 37

S00847

Ig gamma-2c chain C region - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C:Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 23-Jul-1999

C:Accession: S00847

R:Bruggemann, M.; Delmasstro-Galfre, P.; Waldmann, H.; Calabi, F.

Eur. J. Immunol. 18, 317-319, 1988

A:Title: Sequence of a rat immunoglobulin gamma-2c heavy chain constant region cDNA: ext

A:Reference number: S00847; MUID:88166903; PMID:3127222

A:Accession: S00847

A:Molecule type: mRNA

A:Residues: 1-329 &lt;BRU&gt;

A:Cross-references: EMBL:X07189; NID:957602; PIDN:CAA30169.1; PID:g663228



C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: immunoglobulin  
F:20-84/Domain: immunoglobulin homology <IM>

Query Match 23.9%; Score 817.5; DB 2; Length 329;

Best Local Similarity 55.9%; Pred. No. 2,8e-41; Matches 152; Conservative 46; Mismatches 51; Indels 23; Gaps 4;

```

QY 375 SGQVLLSNIKY-LPTWSTPVPKSCDKTHTC-----PP----CPAEL 413
DB 59 SGLYLTSSSVYPSSTMS--QTVCSVAHPATKSNLKRIEPRPKRPDPIDCSODN 116
QY 414 LGGSEVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYDGVENNAKTPRE 473
DB 117 LGRPSVFIFPPKPKDILMITLTPKTCVVVDVSEEPDVSFWDNRYVFTAQTPHE 176
QY 474 QVNSRYRVSVLTVLHODMLNGEKYCKVSNKALPAPKEITISKAKGPPEOVYTLPPS 533
DB 177 QLNGTFRRVSTLHIOHDMGSKERCKVNNKDLPSLEKTSKPRGAKTPQVTTIPP 236
QY 534 RDELTKNOVSLTCLVKGFPSPDIIVWESNGQPENNYKTPPVLDSDGSPFLYSKLTVDK 593
DB 237 RQGNKKNVSLTCWTSFYSPASISVEMRNGELBDYKNTLPVLDSDGSPFLYSKLTVD 296
QY 594 SRWQGNVPSGVMEHALNHTYKSLSPG 625
DB 297 DSMWRGDIYTCVVAHEALHNHTYKSLSPG 328

```

## RESULT 38

G2MSA  
Ig gamma-2a chain C region, secreted form (allele b) - mouse

C:Species: Mus musculus (house mouse)

C:Date: 02-Apr-1982 #sequence\_revision 02-Apr-1982 #text\_change 16-Jul-1999

C:Accession: A02153; A32656

R:Schreier, P.H.; Botwell, A.L.M.; Mueller-Hill, B.; Baltimore, D.

Proc. Natl. Acad. Sci. U.S.A. 78: 4495-4499, 1981

A:Title: Multiple differences between the nucleic acid sequences of the IgG2a(a) and IgG

A:Reference number: A02153; MUID:82037861; PMID:6170065

A:Accession: A02153

A:Molecule type: mRNA

A:Residues: 1-335 <SCH>

A:Cross-references: GB:J00479

A:Experimental source: strain C57BL/6

R:Dognin, M.J.; Lauwereys, M.; Strohberg, A.D.

Proc. Natl. Acad. Sci. U.S.A. 78: 4031-4035, 1981

A:Title: Multiple amino acid substitutions between murine gamma 2a heavy chain Fc region

A:Reference number: A32656; MUID:82037777; PMID:6794027

A:Accession: A32656

A:Molecule type: Protein

A:Residues: 118-267, 'E', 269-328, 'G', 330-334 <DOG>

C:Comment: Lys-335 is removed posttranslationally.

C:Comment: The sequence differs from that of the allele, from BALB/c mice, at 15% of the

C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa)

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglobulin

F:20-84/Domain: immunoglobulin homology <IM1>

F:98-118/Region: hinge

F:142-211/Domain: immunoglobulin homology <IM2>

F:248-315/Domain: immunoglobulin homology <IM3>

F:15/Disulfide bonds: interchain (to light chain) #status predicted

F:27-82,149-209,255-313/Disulfide bonds: #status predicted

F:108,117/Disulfide bonds: interchain (to heavy chain) #status predicted

F:185/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 23.9%; Score 816; DB 1; Length 335;

Best Local Similarity 45.7%; Pred. No. 3.5e-41;

Matches 158; Conservative 52; Mismatches 86; Indels 50; Gaps 3;

```

QY 285 PQALPQVAGSGNTLALAA-----KTGKLHGVNLVWRARQQLKLCFVWGPSPKLM 339
DB 34 PEPVTLTNWNSGLSSGHTPFPALLOSGLYTSSTVTSNTWPSQTITCNVAHPAS---- 89

```

QY 340 LSLKLENKEAKVSKREKPVVNLPEAGMOCLSDSGVLLSNIKVLPTWSTPVEPKSC 399

DB 90 -----STKVDKIEP-----RVITQNPC 108

QY 400 DKHTCPCCAPPELLGGPSVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYD 459

DB 109 PPHQRPVCAAPDLLGGPSVFIPEPKIKDVLMTSLSMVTCVVVDVSEDPDQVISMFWN 168

QY 460 GVEYNAKTPREQRYSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPKEITISKAK 519

DB 169 NVEVHTAQOTHRNDVNSTLRVSALEPIQODMWSGKFKVNNRLPSPIETISKPR 228

QY 520 GQPREPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDIIVWESNGQPENNYKTPPVLD 579

DB 229 GPVRAPOVYVLPRAEEMTKKESLTCTMIGFLPAEIAVDWTSNGRTEQVYKNTATVLD 288

QY 580 DGSFPLYSKLTVDKSRWQGNVPSGVMEHALNHTYKSLSPG 625

DB 289 DGSYFMYSKLRVOKSTWERSLFCASVVAHEVLHNHLLTKTISRSLG 334

## RESULT 39

G2MSA

Ig gamma-2a chain C region, secreted form (allele a) - mouse

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1980 #sequence\_revision 01-Sep-1981 #text\_change 16-Jun-2000

C:Accession: A02152; A32657; A32658

R:Silkorn, J.L.; Aufrey, C.; Rougeon, F.

Nucleic Acids Res. 8: 3143-3155, 1980

A:Title: Structure of the constant and 3' untranslated regions of the murine Balb/c gamma

A:Reference number: A02152; MUID:81076554; PMID:6777755

A:Accession: A02152

A:Molecule type: mRNA

A:Residues: 1-330 <STK>

A:Cross-references: GB:V00798; NID:G51835; PIDN:CAA24178.1; PID:G1333984

R:Yamawaki-Kataoka, Y.; Miyata, T.; Honjo, T.

Nucleic Acids Res. 9: 1365-1381, 1981

A:Title: The complete nucleotide sequence of mouse immunoglobulin gamma 2 A gene and evo]

A:Reference number: A32657; MUID:81198976; PMID:6262729

A:Accession: A32657

A:Molecule type: DNA

A:Residues: 1-330 <YAM>

A:Cross-references: GB:J00470

A:Note: the sequence was determined from the germline gene

R:Ollo, R.; Aufrey, C.; Rougeon, F.

Proc. Natl. Acad. Sci. U.S.A. 78: 2442-2446, 1981

A:Title: Comparison of mouse immunoglobulin gamma-2A and gamma-2B chain genes suggests tr

A:Reference number: A32658; MUID:81223894; PMID:6787604

A:Accession: A32658

A:Molecule type: DNA

A:Residues: 1-330 <OLL>

A:Note: the sequence was determined from the germline gene

R:Bourgois, A.; Fougeron, M.; Rocca-Serra, J.

Eur. J. Biochem. 43: 423-435, 1974

A:Title: Determination of the primary structure of a mouse IgG2a immunoglobulin: amino-ac

A:Reference number: A32659; MUID:74175517; PMID:4811970

A:Contents: annotation; myeloma protein MOPC 173

A:Note: this is one paper in a series reporting the sequence; for additional references,

R:de Preval, C.; Fougeron, M.

Eur. J. Biochem. 30: 452-462, 1972

A:Title: Determination of the primary structure of a mouse gammaG2a immunoglobulin. Ident

A:Reference number: A32660; MUID:73056887; PMID:4555406

A:Contents: annotation; MOPC 173, disulfide bonds

C:Genetics:

A:Introns: 1/1; 98/1; 114/1; 224/1

C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa)

C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglobulin

F:20-84/Domain: immunoglobulin homology <IM1>











```

RESULT 43
S01321
Ig gamma-2b chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 23-Jul-1999
C:Accession: S01321
R:de Waele, P.; Feyers, V.; van de Voorde, A.; Molemans, F.; Fiers, W.
Eur. J. Biochem. 176, 287-295, 1988
A:Title: Expression in non-lymphoid cells of mouse recombinant immunoglobulin directed a
A:Reference number: S01320; MUID:86329081; PMID:3138116
A:Accession: S01321
A:Molecule type: mRNA
A:Residues: 1-475 <DB1>
A:Cross-references: EMBL:X13188; NID:951780; PIDN:CAA31580.1; PID:951781
A>Note: this sequence was determined from the differentiated gene
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-475/Product: Ig gamma-2b chain #status predicted <MAT>
F:159-223/Domain: immunoglobulin homology <IMM>

Query Match      22.9%; Score 782.5; DB 2; Length 475;
Best Local Similarity 33.8%; Pred. No. 5.2e-39;
Matches 205; Conservative 67; Mismatches 162; Indels 173; Gaps 17;

QY 30 LGKGGPVELTCTAS--QKKSIGFWKNSNQIKLGNQ-----SFLTKGPEKLNDRAD 81
DB 30 LARPGASVKSICASGYTLTSYISN-----VKORTGQLEWIGETIPSGNSYFEKKR 84
QY 82 SRRSLW---DQGNFPIIKNLKIEDSDYICEVEDKEVEQLVFLGTANSDTHLQOGS 138
DB 85 GKATLTVDSKSSSTAYVHLISLTSSEDAVFC---GPRQGLLPFG-----YWGQG 132
QY 139 LITLESPPSSSSVQCRSPKKNIGGKTLVSQLELDGSGTWCTVQNKQKVEFKID 198
DB 133 TLVTASAAKTPPSVYPLAPGCGDTG----- 159
QY 199 IYVLARQKASIIYKKEGEVEPSPPLAFTVEKLTGSGELMGAERASSKSWITFDLKN 258
DB 160 -----SSVTL---GCLVKGYPFESVTVT-----WNSGLSS----- 188
QY 259 KEYSVKRVYQDPRLQNGKLLPLHLTPQALPOYAGSNTLALAEATGKLHDEVLVNR 318
DB 189 -----VH-TFPALL-----QSGLYTMSSTVTPS 211
QY 319 ATOLQNLTCFVNGPSPKMLSLKLENKAKVSKREKPVWLINPEAGMOCILSDSGOV 378
DB 212 STWPSQTVTCVNAHPAS-----STYVDKLEP-----SG-- 240
QY 379 LLESNTKVLPTWSTPVPEPKSCDKTHTCPPCAPELLGSPSVLPFPKPKDTLMISRTPEV 438
DB 241 -----PT--STINP--CPRECKCHKCPAPNLGSPSVLPFPNITVDVIMTISTLPV 287
QY 439 TCVVVDVSHEDPEVKENMYVDGVEVNAHNAKTKPREEOYNSTYRVSVLTJLHODMLNGKEY 498
DB 288 TCVVVDVSEDDPVGQSWFNNVEVTLAQOTHREDYNTIRVSLALPQHODMNGKGF 347
QY 499 KCVNSKALPAPLEKTIKAKAGQPREPOVYTLPPSDELTKNQVSLTCLVKGFPSPDIAY 558
DB 348 KCVNNKDLPAPIERTISKIKGIVRAPOVYIISPPPEQLSRKQVSLTCLAVGSPSPDISV 407
QY 559 EMESNGQPENNYKTPPVLDSDGSFLYSKLTJVDKSRMGQGVNFCVSWHEALHNHYTKK 618
DB 408 EMTSNQHTENYKDTAPVLDSDGSFYIYSKLNKTKSKMEKTDSEFCVNRHEGLKNYYLKK 467
QY 619 SLSLSPG 625
DB 468 TISRSPG 474

```

```

RESULT 44
S06611
Ig gamma-2 chain C region (clone 32.2) - bovine (fragment)

```

```

C:Species: Bos primigenius taurus (cattle)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 16-Jul-1999
C:Accession: S06611; B11303
R:Symons, D.B.A.; Clarkson, C.A.; Beale, D.
Mol. Immunol. 26, 841-850, 1989
A:Title: Structure of bovine immunoglobulin constant region heavy chain gamma 1 and gamma
A:Reference number: S06610; MUID:90097956; PMID:2513487
A:Accession: S06611
A:Molecule type: DNA
A:Residues: 1-327 <SYM>
A:Cross-references: EMBL:X16702
A>Note: the sequence was determined from the germline gene
C:Genetics:
A:Gene: Ig CH gamma 2
A:Introns: 99/1; 112/1; 219/1
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: glycoprotein; immunoglobulin; membrane protein
F:20-85/Domain: immunoglobulin homology <IMM>
F:175/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match      22.7%; Score 774; DB 2; Length 327;
Best Local Similarity 54.4%; Pred. No. 1e-38;
Matches 148; Conservative 42; Mismatches 64; Indels 18; Gaps 3;

QY 371 LLDSDGOVLLLESNIKVLPTWSTPVPEKSCDKTHTCPPCAPELLG----- 415
DB 56 VIQSSGLYSLSMTYV-PASSGTYOTFCVNAHPASSTRKVDKAVSSDCSPNNQHCVR 114
QY 416 GPSVFLPPEKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAHNAKTKPREQY 475
DB 115 EPSVFIFPPEKPKDTLMITGTEPTEVCVVVNVGHNDPEVQFSWFDVDEVHTATKREDEQ 174
QY 476 NSTYRVSVLYTLVHDDMLNGKEYKKKQVSKALPAPIEKTISKAKQPREPOVYTLPPSRD 535
DB 175 NSTYRVSVLPIQHODMNGKFKKCKVNIKGLSASIVAIISRSKQPAPEPOVYVLDPPKE 234
QY 536 ELTKQVSLTCLVKGFPSPDIAYEVESNGQ--PENNYKTPPVLDSDGSFLYSKLTJVDK 593
DB 235 ELKSTVSLTCLVKGFPSPDIAYEVESNGQ--PENNYKTPPVLDSDGSFLYSKLTJVDK 294
QY 594 SRMQQGVNFCVSWHEALHNHYTKSLSPG 625
DB 295 NSMGRGDTYTCVNMHEALHNHYTKSKSAG 326

RESULT 45
146732
Ig gamma heavy chain constant region - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 21-Jan-2000
C:Accession: 146732
R:Heidmann, O.; Rougeon, F.
Nucleic Acids Res. 10, 1535-1545, 1982
A:Title: Molecular cloning of rabbit gamma heavy chain mRNA.
A:Reference number: 146732; MUID:82174328; PMID:6280149
A:Accession: 146732
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-180 <HEI>
A:Cross-references: GB:J00665; NID:9165109; PIDN:AAA31288.1; PID:9165110
C:Superfamily: immunoglobulin C region; immunoglobulin homology
F:93-160/Domain: immunoglobulin homology <IMM>

Query Match      20.6%; Score 702; DB 2; Length 180;
Best Local Similarity 69.7%; Pred. No. 8.7e-35;
Matches 124; Conservative 26; Mismatches 28; Indels 0; Gaps 0;

QY 448 EDPVKFNMYVDGVEVNAHNAKTKPREOYNSTYRVSVLTJLHODMLNGKEYKKCVSNAL 507
DB 2 DDPVQFTWYINNEQVTRAPPLKQPNSTIRVSTLPIAHQDWLKRKEKCKVHNVAL 61
QY 508 PAPIEKTISKAKQPREPOVYTLPPSDELTKNQVSLTCLVKGFPSPDIAYEVESNGQPE 567

```



[illegible]

## RESULT 46

IG heavy chain V-III region (ART) - human (fragments)  
C:Species: Homo sapiens (man)  
C:Date: 16-Nov-1990 #sequence\_revision 16-Nov-1990 #text\_change 21-Jan-2000  
C:Accession: A36040  
R:Eultz, M.; Weiss, D.T.; Solomon, A.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6542-6546, 1990  
A:Title: Immunoglobulin heavy-chain-associated amyloidosis.  
A:Reference number: A36040; PMID:90370821; PMID:2118650  
A:Accession: A36040  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-218 <EUL>  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
C:/132-199/Domain: immunoglobulin homology <IMM>

Query Match	16.8%;	Score 574.5;	DB 2;	Length 218;
Best Local Similarity	77.5%;	Pred. No. 3.9e-27;		
Matches 110; Conservative	3;	Mismatches 12;	Indels 17;	Gaps 1

484 VLTVLHQDWLNGKEYKCKVSNKALPAPIEXTISAKGQPREPOVYTTLPSSRDELTKNQVS 543

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

100

[illegible]

Db 197 CSVMHEGLHNHYTOKSLSPG 218

## RESULT 47

Ig heavy chain VHIII-D-JH-CH3 region - human  
C/Species: Homo sapiens (man)  
C/Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 21-Jan-2000  
C/Accession: S69340  
R/Kamlichl, A.A.; Auncouturier, P.; Preud'homme, J.L.; Cogne, M.  
Eur. J. Biochem. 229, 54-60, 1995  
A/Title: Structure of abnormal heavy chains in human heavy-chain-deposition disease  
A/Reference number: S69339; MUID:95262687; PMID:7744049  
A/Accession: S69340  
A/Status: Preliminary  
A/Molecule type: mRNA  
A/Residues: 1-249 <KHA>  
A/CrossB-references: EMBL:X81596  
A/Superfamily: Immunoglobulin C region; immunoglobulin homology  
A/162-229/Domain: immunoglobulin homology <IMM>

Query Match	16.8%;	Score 572.5;	DB 2;	Length 249;
Best Local Similarity	62.9%;	Pred. No. 6.1e-27;		
Matches 117; Conservative	9;	Mismatches 23;	Indels 37;	Gaps 3

460 GVEVHNAKTKPREEQYNSTYRVSVLT-----VLHQDWLNGKEYKCKVSN 504

DD 80 GDSVAGKF I ISKNSEENVU I UQWNSLAEIAV I I CANNVGSV IN----- 129

[illegible]

**THE UNIVERSITY OF CHICAGO**

Db 183 WESNGPENNKTTPVLDSDGSFFLYSKLTVDXSRRMOGNVFCSCVMEALHNHYTKS 242

QY 620 LSLSPG 625

Db 243 LSLSPG 248

## RESULT 48

Ig gamma-1 chain C region (15C5) - mouse (fragment)  
 C/Species: Mus musculus (house mouse)  
 C/Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #next\_change 23-Jul-1999  
 C/Accession: S14236  
 R/Vandamme, A.M.; Bulens, F.; Benrux, H.; Nelles, L.; Lijnen, R.H.; Collen, D.  
 Eur. J. Biochem. 192, 767-775, 1990  
 A/Title: Construction and characterization of a recombinant murine monoclonal antibody d  
 A/Reference number: S14236; MUID:91006173; PMID:2209622  
 A/Molecule type: mRNA  
 A/Residues: 1-152 <EVB>  
 A/Cross-references: EMBL:X536393; NID:gs51617; PIDD:CA339804.1; PIDD:gs51618  
 C/Superfamily: Immunoglobulin C region; Immunoglobulin homology  
 C/Keywords: Immunoglobulin  
 C/65-132/Domain: immunoglobulin homology <IMM>

Query Match	16.6%	Score 566;	DB 2;	length 152;
Best Local Similarity	61.6%	Pred. No. 7.9e-27;		
Matches 93;	Conservative 33;	Mismatches 25;	Indels 0;	Gaps 0

475 YNSTYRVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPVYTLPPSR 534

22 A T N O I F N O V O E E F I T H I Q U O N E I N S I G N I F I C A N T V A R I A N C E S W A S F O U N D I N T H E I D E N T I D O N A N N O U N C E I N T H E Q U A L I T Y A T T E S T A T I O N

[illegible]

FOR PROSECUTION AND DEFENSE

Db 121 NWEAGNTFTCSVLHEGLHNHHTKSLSHSPG 151

## RESULT 49

B46529  
 Ig Y heavy chain (7.8S) - duck  
 N:Alternate names: Ig gamma chain (7.8S)  
 C:Species: *Anas platyrhynchos* (domestic duck)  
 C:Date: 18-Jun-1993 #sequence\_revision 18-Nov-1994 #ext\_change 21-Jul-2000  
 C:Accession: B46529; S20759  
 R:Magor, K.E.; Warr, G.W.; Middleton, D.; Wilson, M.R.; Higgins, D.A.  
 J. Immunol. 149, 2627-2633, 1992  
 A:Title: Structural relationship between the two IgY of the duck, *Anas platyrhynchos*: molecular cloning, sequencing and expression of the heavy chain genes  
 A:Reference number: A46529; WUID:93017865; PMID:1401901  
 A:Accession: B46529  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-572 <MAG>  
 A:Cross-references: EMBL:X65219; NID:g62442; PIDD:CAA46322.1; PID:g62443  
 A:Experimental source: spleen  
 A:Note: sequence extracted from NCBI backbone (NCBI:116127)  
 C:Superfamily: Immunoglobulin C region; Immunoglobulin homology  
 C:Keywords: immunoglobulin  
 C:37-130/Domain: immunoglobulin homology <IMM>

Query Match	11.6%;	Score 394.5;	DB 2;	Length 572;
Best Local Similarity	26.3%;	Pred. No. 6.1e-16;		
Matches 141; Conservative	80;	Mismatches 207;	Indels 109;	Gaps 22

150 SPSPVQCRSPRKNI-QGKTLVSQLELODSTWTCTVLONQKKEFKIDIWLAQKAS 208

03 AFHVNAGNF ILSKNNQDS IMLQMNLSLNAED I M I I C A-----KSS 123



```

QY 209 SIYKKEGQVEFS-----FPLAFVEKLTGS-----GELMWQAEKASS 247
DB 130 DIDLMHGTEVAASSSGPTAPSVFPISSCGSGTQOQPVWGLATGYIPBPVTFPSMGASG 189
QY 248 SKSMITFDLKNKEVSKRVTQDPKLOMGKLPDLTLPOALPOYAGSNL---TLALEAK 304
DB 190 AYS-----VTVPETHGVGP-----HKRASFLRPRHAGAGPFTCSVNHQAT 230
QY 305 TGTLDHOEVLVYWMRA-----TQLOKNLTCEVWGFPSPKLMLSLKLENKAVSKREKPV 358
DB 231 KRSLTQNVGECVAGGEPPTPEPVQVHLHSVCSLTIGDDSVELCVI-----TGFSPPV 282
QY 359 ---WVLNPPAGMWQCLSDSGVLESNIKV-LPTWSTPV--EPKSCDTHTC---PP- 407
DB 283 EYEWLVD---GAPHLVATMTTRPQRAGSKTYVATSGTNVSRDMAGAFCTRVHGPRT 339
QY 408 -----CPABELLGGS-----VFLPPPKQDTLMSRTPEVTCVVVDVSHEDPEV 452
DB 340 GGTAGQGHARFCPS---GAQSCSPQIFVVPSP--GLYIRQDAKVCVLVLP--SDASL 394
QY 453 KFNWYVDGVEVHNNAKTPREEOYNSTRYVSVLTVLHODMLNGKEKCKVSNKALPAPIE 512
DB 395 SLSWTRKSGALRPDPMLTEHNGTFYASSSLAISTQDMLAGERFTCTVQHEDELFPVIG 454
QY 513 KTIKAKGQPREPOVYTLTPPSRDELTKNQVSLTCLVKGFPSPDIAYEM--ESNGQEPNNY 570
DB 455 KSLAKAGAVTAAYITTFPPHAEELSLAEVTLTCLVRGTPQPEHVEQVRLRNNSVPAALF 514
QY 571 KTTTPVLD--SDGSFFLYSKLTVDKSRWQGVNFSQSVNHEALHNHYTKSLSLSPG 625
DB 515 VTPPLKEPNQDGTFFLYSKMTVPKASWQGVSYACMVVHEGLPMKFTQRLQKTPG 571

```

## RESULT 50

S14683  
Ig mu chain precursor, membrane-bound (clone 201) - human  
C1:Species: Homo sapiens (man)  
C1:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 23-Jul-1999  
C1:Accession: S14683; S08047  
R1:Friedlander, R.M.; Nusse, M.C.; Leder, P.  
Nucleic Acids Res. 18, 4278, 1990  
A1:Title: Complete nucleotide sequence of the membrane form of the human IgM heavy chain.  
A1:Reference number: S14683; MUID:90332450; PMID:2115996  
A1:Accession: S14683  
A1:Molecule type: mRNA  
A1:Residues: 1-627 <FRT>  
A1:Cross-references: EMBL:X17115; NID:G33450; PID:CAA34971.1; PID:G33451  
C1:Superfamily: Immunoglobulin C region; immunoglobulin homology  
C1:Keywords: immunoglobulin; membrane protein  
F1:1-15/Domain: signal sequence #status predicted <Sig>  
F1:16-627/Product: Ig mu chain #status predicted <MAT>  
F1:34-117/Domain: immunoglobulin homology <IMM>

```

Query Match      11.4%; Score 390.5; DB 2; Length 627;
Best Local Similarity 22.6%; Pred. No. 1.2e-15;
Matches 157; Conservative 103; Mismatches 231; Indels 205; Gaps 30;

QY 32 KKGDTVELTCTAS-----QKSIQFHKNNSQIKILNQSGSFLTKGPKSLND 78
DB 32 KFGSSVYKSCKASGCTFSSVSAISWVQAEGGLEMM--GGILPIFGT-ANYAQKFGQRTVI 89
QY 79 RADSRSLWDQGNFPLIKNLKIEDSDTYICEVEDKEVOLL---VFGLTANSDPHL-- 133
DB 90 TADESTS-----TAYELSSLRSEDTAYYYC-----AKTGILGPYSSGWYPSNDYYYG 138
QY 134 --LOGSLTLTLESPPGSSPS---VQCR----- 156
DB 139 MDVWGCTTVTVSSGSASAPTLFPLVSCNSPBDTSSVAVGLAQDPLPDSITFSWKYN 198
QY 157 -----SPRG--KNIOGKTLVSQLELD---SGT---WTCTVLONQKKEPKIDIVVL 202
DB 199 NSDISSTRGFPSVLRGKVAATSOVLPSKDVWQGTDEHVVCKVQHPNKNKKNVPLPVI 258

```

```

QY 203 AF--OKASSIYKKEGQVEFSFPLAFVEKLTGSGELMWQAEKASSK---SWITFDLK 257
DB 259 AELPPKSVFVPPRRG---FQGNP-----RSKSKLICQATGSPRQIQVSW----- 301
QY 258 NKEVSKRVTQDPKLOMGKLPDLTLPOALPOYAGSNLTLAEAKTGKLDHOEVLVYM 317
DB 302 -----LREKQV-----GSGVTTIDQVQAEAKES----- 324
QY 318 RATQLOKNLTCEVWGFPSPKLMLSLKLENKAVSKREKPVVVLNPPAGMWQCLSDSGQ 377
DB 325 -----GPTTYKVTSTLTKESD-----WL---SQSMFTCRDRHGL 357
QY 378 VLE--SNIKVPTWSTPVPEKSCDTHTCPPCAPABELLGSSVFLPPPKQDTLMSRT 436
DB 358 TFGQNASMCVDDQTAIR-----VFALPPS--FASIFLTST 393
QY 437 EYTCVVVDVSHEDPEVKFNWYVDGVEVHNNAKTPREEOYNSTRYVSVLTVLHODMLNG 496
DB 394 KLTCLVTLDTYD--SVTISWTRQNGEAVKTHNTNISESHPNATFSAVGBASICEDDMNSGE 452
QY 497 EYKCKVSNKALPAPIEKTIKAKGQPREPOVYTLTPPSRDELT--KNQVSLTCLVKGFP 554
DB 453 RFTCTVTHTDLPSPKQTIISPKGVALLRPDVYLLPPAREQNLRESATITCTLVGFS 512
QY 555 DIAVWESNGQPP--ENNKTTTPVLD--SDGSFFLYSKLTVDKSRWQGVNFSQSVNHEA 610
DB 513 DVFVOMQGRQDPLSEKTYTSLAPMPEPOAPRGYFNSHLITYSEEMNNGETTYTCVAAHA 572
QY 611 LHNHYTKSLSLSPGLQDDETCAEQDDELGLMTT 646
DB 573 LPNRYTERTVDS-----TEGEVSADIEGFENLWAT 603

```

## RESULT 51

S04845  
Ig heavy chain precursor - African clawed frog (fragment)  
C1:Species: Xenopus laevis (African clawed frog)  
C1:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 26-Aug-1999  
C1:Accession: S04845; S05695  
R1:Amemiya, C.T.; Haire, R.N.; Litman, G.W.  
Nucleic Acids Res. 17, 5388, 1989  
A1:Title: Nucleotide sequence of a cDNA encoding a third distinct Xenopus immunoglobulin  
A1:Reference number: S04845; MUID:89345103; PMID:2503814  
A1:Accession: S04845  
A1:Molecule type: mRNA  
A1:Residues: 1-549 <AME>  
A1:Cross-references: EMBL:X15114  
A1:Submitted to the EMBL Data Library, April 1989  
A1:Reference number: S05695  
A1:Accession: S05695  
A1:Molecule type: mRNA  
A1:Residues: 1-549 <AME>  
A1:Cross-references: EMBL:X15114; NID:G64799; PID:9763031  
C1:Superfamily: Immunoglobulin C region; immunoglobulin homology  
C1:Keywords: glycoprotein; heterotetramer; immunoglobulin  
F1:26-109/Domain: immunoglobulin homology <IMM>  
F1:281-294/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

Query Match      11.4%; Score 389; DB 2; Length 549;
Best Local Similarity 25.6%; Pred. No. 1.2e-15;
Matches 153; Conservative 82; Mismatches 211; Indels 152; Gaps 25;

QY 110 EYEDQKEEVL--LVFGILT-ANSDTHLQ---QGSLLTLESPPGSSPSVQCRSPRGK-- 161
DB 21 EIKSGESIKLSCKTSGYTFNWIHWIQVPGKQLQIGRIYPEDADTDVSSSYQGRH 80
QY 162 ---NIQGKTLVSQLELDPSGTCTVVLQNKQKVEKIDIVLAFAKASSIYKKEGE 217
DB 81 ISTDPSQSTTFPLQNLNKLVEDTAIYYCA-----REGV 112
QY 218 QVEFSF-----PLAFVEKLTGSGELMWQAEKASSKSWITFDLKNKEV 261

```



```

Db      113 GVEFDYWGQWTVTSATLHAPSVPLRPCCGS-----SSDSHTTIGCLSTGF 162
Oy      262 SVKRVTOPKLOMGKKLPLHLTLPOALPO---YAGSGNLTALAEKTKLHOEVNLV-VM 317
Db      163 LPAPV--DYKNNSSGTTGKLPFAVLQOOSGLFASSQTLTPLSMWAKKSEFCEVHEKP 220
Oy      318 RATLOQKULTCG-----VWGPTSPK-----LMSLKLKNEKAKVSKREKPVV 360
Db      221 TSTKTOKIECQDEDEPIEPTVEIILQGPCASSKVELLCLITGVAPSEIKYH-----WL 274
Oy      361 LNPBAGMMOCLSDSGQV-----LLESNIKVLPTWSPVPEPKSCDKTHT 404
Db      275 LN-----GQVLTNISPNSKPCKEENGTFSSRSKV---SVPKEDMNSDSYT 317
Oy      405 CPPC-PA-----PELLGPPSVFLFPPEPKDTLMSRTEPVYCVVDVSHEDPE 451
Db      318 CKNTHPAHHTKEASTKKCDERTATPKYDVLPRPKD-LVLYKAKVYCVISRASTP-D 375
Oy      452 VKFNN-VYDGVFNHNAKTKPREEQVNSTYRVVSVLTVLHQMVLNGEKYCKVSNKALPAP 510
Db      376 LTVQMSRSDGKALAPDSAP-EKAYDGFPTVSKTLKISPGDWENKQCNCKVHPDLRSP 434
Oy      511 IEKTSKAKGQREPOVYTLTPSRDELTKNOVSLTCLYKGFPSDIANWESNG--OPEN 568
Db      435 IEKTSKQSDPGETEPTTLTLLPPSDDELKNDFTSLICMLKNRPQDIYVFWKXDGVTLEDE 494
Oy      569 NYKTPPVLDSDGSFPL-YSKLTVDKSRMOQGNVFCSCVMHEALHNHYTKSLSPG 625
Db      495 YMTTTPVLEEEBEFISFSKLTIRASDMKGAATVSCI---MAINTISQADIKKNG 548

```

## RESULT 52

MHSM

IG mu chain C region, membrane-bound form - mouse

```

C:Species: Mus musculus (house mouse)
C>Date: 31-Oct-1980 #sequence_revision 30-Jun-1991 #text_change 22-Jun-1999
C:Accession: A02167; A37517; B02166
R:Early, P.; Rogers, J.; Davis, M.; Calame, K.; Bond, M.; Wall, R.; Hood, L.
Cell 20, 313-319, 1980
A:Title: Two mRNAs can be produced from a single immunoglobulin mu gene by alternative R
A:Reference number: A02167; MUID:80222874; PMID:6771020
A:Accession: A02167
A:Molecule type: DNA
A:Residues: 436-476 <BAR>
A:Cross-references: GB:V00816; GB:J00444; NID:952343; PIDN:CAA24197.1; PID:952344
R:Rogers, J.; Early, P.; Carter, C.; Calame, K.; Bond, M.; Hood, L.; Wall, R.
Cell 20, 303-312, 1980
A:Title: Two mRNAs with different 3' ends encode membrane-bound and secreted forms of in
A:Reference number: A37517; MUID:80222873; PMID:6771019
A:Contents: MOPC 104E
A:Accession: A37517
A:Molecule type: mRNA
A:Residues: 410-476 <ROG>
A:Cross-references: GB:V00821; NID:952355; PIDN:CAA24202.1; PID:9817972
R:Kawakami, T.; Takahashi, N.; Honjo, T.
Nucleic Acids Res. 8, 3933-3945, 1980
A:Title: Complete nucleotide sequence of mouse immunoglobulin mu gene and comparison with
A:Reference number: A02166; MUID:81076590; PMID:6255422
A:Accession: B02166
A:Molecule type: DNA
A:Residues: 1-435; GKPTLVNVSIMSDTGCTC' <KAN>
C:Comment: The sequence of residues 1-409 was assumed to be identical with the correspo
C:Comment: During differentiation, B lymphocytes switch from expression of membrane-bou
8.
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (ka
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 18
C:Superfamily: Immunoglobulin C region; Immunoglobulin homology
C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobu
P.21-91/Domain: immunoglobulin homology <IMM1>
P.129-201/Domain: immunoglobulin homology <IMM2>
F.239-307/Domain: immunoglobulin homology <IMM3>
F.346-417/Domain: immunoglobulin homology <IMM4>
F.436-476/Domain: carboxyl-terminal <CTS>

```

F.14/Disulfide bonds: Interchain (to light chain) #status predicted  
 F.28-89,136-159,246-305,353-415/Disulfide bonds: #status predicted  
 F.46,211,243,258,281/Binding site: carbohydrate (asn) (covalent) #status predicted  
 F.216/Disulfide bonds: Interchain (to heavy chain) #status predicted  
 F.293/Disulfide bonds: Interchain (to mu chain in another subunit) #status predicted

Query Match 11.0%; Score 375.5; DB 1; Length 476;  
 Best Local Similarity 22.9%; Pred. No. 6,4e-15;  
 Matches 127; Conservative 83; Mismatches 188; Indels 157; Gaps 18;

```

Oy      144 ESPGSSPSVQGNRP-RGNIIQGGKTLVSQLELQDSGTCTYVLQOKVYEFKIDIVL 202
Db      3 QSFVNVPLVSCSPSPDKMLVAMGCLARDLFTSTSTFN----- 43
Oy      203 AFQASSIVYKKEGEQVEFSPFLAFVTEKLTGSGELMWQABASSSKSWITFDLKNKVS 262
Db      44 -----YQNTVEVQ-----GRTFTPLRGKGYLA 68
Oy      263 VKRVTOPKLOM-----GKKLPLHLTLPOALPOYAGSGNLTALAEKTKL 308
Db      69 TSGVLSPKSILREGDEYLVCKIHGGKNDLHVPI-----AVAEW 110
Oy      309 HOEVNLV-----MRATLOQKULTCGVWGPSPKMLSLKLEKAKVSKREKPVV 362
Db      111 NPVNVVFPVRPDGFGSPAPRKSKLICATNFTPKPIVSWLKQKGLVESGFTTDPVTIEN 170
Oy      363 ---PEA-----GMWQCLSDSGQVLTLSNLIKVLPTWSTPVEPKSCDK 401
Db      171 KGSFPQYKYITSLTISEIDMLNLYTTCVDRHGLTFLK-----NV 212
Oy      402 THTCPPEABELGSPVFLFPPEPKDTLMSRTEPVYCVVDVSHEDPEVKNVYDV 461
Db      213 SSTCAASPSYDIL---TFTIPSPFAD-IFLSKANLTCLVSNLATE-FLNISWASQSG 266
Oy      462 EVNNAKTKPREEQVNSTYRVVSVLTVLHQMVLNGEKYCKVSNKALPAPLEKTSKAKQ 521
Db      267 EPLETKIKIMESHNGTFSAKGVASVCEVDMNRKEVCTVTHRDLPSPQKFTSK--- 322
Oy      522 PRE-----POVYTLPPSRDELTKNOVSLTCLYKGFPSDIANWESNGO--PENNYKT 573
Db      323 PNEHNGKRPAYLLPPAREQNLARESATVYCLYKGFSPALISVOMLORGQLQDEKVT 382
Oy      574 PVYLD--SDGSFPLYSKLTVDKSRMOQGNVFCSCVMHEALHNHYTKSLSPGLQDET 631
Db      383 APMEPEAPGFYFHTSLITVTEEMNSGETYTCVVGHEALPHVTERTVKSS-----TEG 437
Oy      632 CAAQDGEHLQGLWT 646
Db      438 EVNAEEGFENLWTT 452

```

## RESULT 53

S25705

IG mu chain - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C&gt;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 23-Jul-1999

C:Accession: S25705

R:Patil, S.; Nau, F.

Mol. Immunol. 29, 829-836, 1992

A:Title: Isolation and sequence of a cDNA coding for the immunoglobulin mu chain of the

A:Reference number: S25705; MUID:92342148; PMID:1635560

A:Accession: S25705

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-592 &lt;PAT&gt;

A:Cross-references: EMBL:X5994; NID:g1269; PIDN:CAA42611.1; PID:g1270

C:Superfamily: Immunoglobulin C region; Immunoglobulin homology

C:Keywords: immunoglobulin

F.483-554/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 10.9%; Score 371.5; DB 2; Length 592;  
 Best Local Similarity 22.7%; Pred. No. 1.5e-14;  
 Matches 156; Conservative 96; Mismatches 253; Indels 181; Gaps 29;



```

OY 6 PRRLHLVLQ---ALLPAAIQGNKVVLAGKKGDITELCTASQKSKIQFH--MKNSNIK 60
Db 3 PLMTLTFPLSAPRGVLSOVQLOESGSPSLVKSGSETLSLTCTVSGSGLTVNHNHSMIRQASGK 62
OY 1LGNQGSFLTKGSPSLKLNDRADSRSL---WDQGNFPLIIKNIKIEDSDPTYICEVEDQKEE 117
Db 63 MPEMLGIVEKGNQNTYNNPALKSRLSIARDTSKQVSLSSMAIDTAVYIC----- 114
OY 118 VOLVFLGLTANSDTLLQ-----GQSLTTL---ESPFGSSPSVQC-RSPRGKNIQ--- 164
Db 115 -----ARSAGAYFLADVDVIMGRGLLVTVSSSESHPKVFLVSCVSSPSDEMTVALG 166
OY 165 -----GKTLVSQQL-----ELDQSGMTQC---VLNQKKVEXKIDIV 200
Db 167 CLARDFVNSVSFSKFNKSTVSSERFWTFPEVLADGLMSSQVALHSSSTFGQTDGYL 226
OY 201 VLAFO--KASSIVYKKEGOVEFSPPLAFT---VEKLTSGSELMQAEASSSKSWITF 254
Db 227 VCEVQHPRGDEGVGHKGVPREVEVLSPVVSVFVPCNSLSGNG-----NSKSLIQAT 279
OY 255 DLAKNEVSV-----KRVTDQPKLQMGKPLPLHLTLPOLPOLYAGSGULTALAEKTKL 308
Db 280 DFSPKQISLWFRDCKRIVSD-----ISEGVETVQSSPTTYR 317
OY 309 HOEVNLYVMRATQLOKNLTCEYWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNEAGMW 368
Db 318 AYSVLTITIEREKLSSATTCQV-----ENHKEI----- 345
OY 359 QCLLSDSGOVLLIESNIKVLPTWSTPVEBKSCDKHTTCCPCAPALLGSPVFLPPPKKD 428
Db 346 -----FQKNAS-----SSCDAT--PPSPI-----GVFTIIPSPAD 373
OY 429 TLMISRTPEVTCVVVDVSHEDPEVKFN--YVDG--VEYHNAKTKREREQNSTYKRVSVL 485
Db 374 -IFLTKSAKLSGLVYTNLASVD--GLNISWSHONGKALEHTY---FEHLNDTFSARGEA 427
OY 486 TVLHODMLNGKSKYCKYCNKALPARIETKISKAKG-QPREQVYTLPPSRDELTKNQVS 543
Db 428 SVCSGDMWESGEYTCVTHADLPPEPKSAISIKPKDVAKPKPSYVLPPTTRQLSLREBAS 487
OY 544 LTCLVKGFYPSDIAVEMESNGQP--ENNYKTTTPVLDD--SDGSFELYSKLTVDKSRWQOG 599
Db 488 VTCLLVKGAPAPADVFOVQLQKGEPRVAKSKYVTSPPAPEQDPSAYVHSILTVTEEDWEGK 547
OY 600 NVFGSGVNHGALHNHYTQKSLSLSPG 625
Db 548 EYTCVGVGHEALPHMWTERTYDKSTG 573

RESULT 54
S38864
I9 epsilon chain C region - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-May-2001
C/Accession: S38864
R/Kip, B.; Becker, W.; Schlaak, M.
Submitted to the EMBL Data Library November 1993
A/Description: Combination of a defined specificity and desired isotype by cloning of an
A/Reference number: S38864
A/Accession: S38864
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-548 <KIP>
A/Cross-references: EMBL:Z27397; NID:g416537; PIND:CAA01788.1; PID:g940782
C/Superfamily: Immunoglobulin C region; Immunoglobulin homology
F:J353-421/Domain: immunoglobulin homology <IMM>

Query Match 10.8%; Score 370; DB 2; Length 548;
Best Local Similarity 24.2%; Pred. No.1,6e-14;
Matches 157; Conservative 108; Mismatches 217; Indels 166; Gaps 31;

30 LGKKGDITVELCTAS--QKSKIQFH-----KNSNIKILGNQGSFLTKGSPSLKLN----- 77

```

Db 11 LVKGGSLKLSAASGLTFSSYGMWVRQIPDKRLVAVATISGGTY-ITYPPSVVGRFT 69

Qy 78 -DRADSRSLMDQGNFPLI1KNLKIEDSDTYICEVEDQKEEVLVFGFLANSDDLQ 136

Db 70 ISRONAKNTLY-----LQMSLSKSEDTANYVCARQGVSTWHPFAW-----G 111

Qy 137 QSLTLTLESPPGSSPSYQCRSPRGKNIQCGAKTSLVSQLELQDSG-----TWCTVLQ 188

Db 112 QGLTAVTSAGKTPPSPYVPLPG---SAAQTNSMTVLGCLVKGYPEPVTYVWNSGL- 166

Qy 189 NQKVEPFKIDIVLAFQKASSIYKKKEGEVRSFPLAFVEXKLTSGL-----WMQAE 243

Db 167 -----SSGVH-----TFPVLQSDLYTSSSTVPSSTWPS 198

Qy 244 RASSSKSWITFDLKNKEVSVKRVYODPKLQMGKPLPHLTLPOLPOAGSGNLTALBA 303

Db 199 TVTGN-----VAHPASTVVDKKIVPRD-----CGCK-----PCIVP----- 230

Qy 304 KTGKLBQBNLV-----VMRATOLQNLTCFVGPSPKMLSLKLENKAVSKR 354

Db 231 -----EVSSVFPFPKPKVDVLRSTIQ--LYCFIGHILNDVSVSMDDREIDTTLA 281

Qy 355 EKPPVNLNPEAGMMQCLSDSGV--LLESNIKXVLPMTSTFVEKSDCKHTTCPCAPEL 413

Db 282 QT---VLKEBGK---LASTCSKLNITEQOMSESTFTCKVTSQGVADYLAHTRCPDHE- 334

Qy 414 LGGSPVLPFPKPKDPTLMISRTPEVTCVVVDV-SHEDPEKFN-----WYVDGV 461

Db 335 PRGYITLIPSPLD-LYNGAPKPLCLVVDLSEKGNVNTVNOEKKTSVSAQWY---T 350

Qy 462 EVHNAKTRPBEQYNSYTRVSVLTVLHODLMNGEKYCKVSKNALPAPIEKTISKAKG 521

Db 391 KHN-----NATTSITSLPVAKMIGGYQCIVDHPDPKPIVRSITXPGQ 440

Qy 522 PREQVYTLPPSRBELTKNOVSLTCLVKGYYPSDIAVWESNQOPENN--YKTTPLPDS 579

Db 441 RSAEEVYVFPPEEE-SEDKRTITLCLQNFPPEDISVQWIGDKLINSQHSSTTP-LKS 498

Qy 580 DGS---FFLYSKLTVDKSRMOQGNVFSQWHEALNH-HYTKQSLSS 623

Db 499 NGSKRGFFISRLVAKTLMYTKRKQFTCOVHEHLQKPRLEKTISS 546

RESULT 55

MEMS

Ig mu chain C region, secreted form - mouse

C:Species: Mus musculus (house mouse)

C:Date: 31-Oct-1980 #sequence revision 31-Oct-1980 #text change 16-Aug-1996

C:Accession: A02166; A26239; A26240; B02039

R:Kawakami, T.; Takahashi, N.; Honjo, T.

Nucleic Acids Res. 8, 3933-3945, 1980

A:Title: Complete nucleotide sequence of mouse immunoglobulin mu gene and comparison with

A:Reference number: A02166; MUID:81076590; PMID:6255422

A:Accession: A02166

A:Molecule type: DNA

A:Residues: 1-455 <XAM>

A:Cross-references: GB:J00443

A>Note: the sequence was determined from the germ-line gene

R:Goldberg, G.I.; Vanin, E.F.; Ziolka, A.M.; Blattner, F.R.

Gene 15, 33-42, 1981

A:Title: Sequence of the gene for the constant region of the mu chain of Balb/c mouse im

A:Reference number: A26239; MUID:82051295; PMID:6795090

A:Accession: A26239

A:Molecule type: DNA

A:Residues: 1-455 <GOL>

A>Note: the sequence was determined from the germ-line gene

R:Auffray, C.; Rougeon, F.

Gene 12, 77-86, 1980

A:Title: Nucleotide sequence of a cloned cDNA corresponding to secreted mu chain of mouse

A:Reference number: A26240; MUID:8165562; PMID:6260531

A:Accession: A26240



A:Molecule type: mRNA  
 A:Residues: 1-225, 'N', 227-257, 'S', 259-367, 'K', 369-455 <AUF>  
 R:Kehry, M.; Sibley, C.; Fuhrman, J.; Schilling, J.; Hood, L.E.  
 Proc. Natl. Acad. Sci. U.S.A. 76, 2932-2936, 1979  
 A:Reference number: A26241; MUID:79223904; PMID:111247  
 A:Contents: annotation; MOPC 104E  
 A>Note: this sequence has been revised in reference A02039. Carbohydrate binding sites a  
 R:Kehry, M.R.; Fuhrman, J.S.; Schilling, J.W.; Rogers, J.; Sibley, C.H.; Hood, L.E.  
 Biochemistry 21, 5415-5424, 1982  
 A>Title: Complete amino acid sequence of a mouse mu chain: homology among heavy chain cc  
 A:Reference number: A02039; MUID:83075344; PMID:6816276  
 A:Contents: MOPC 104E  
 A:Accession: B02039  
 A:Molecule type: Protein  
 A:Residues: 1-77, 'N', 79-100, 'Q', 102-225, 'N', 227-257, 'T', 259-367, 'K', 369-455 <KEH>  
 C:Genetics:  
 A:Introns: 1/1, 106/1, 219/1, 325/1  
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap  
 hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 16  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob  
 F:128-91/Domain: immunoglobulin homology <IMM1>  
 F:128-201/Domain: immunoglobulin homology <IMM2>  
 F:239-307/Domain: immunoglobulin homology <IMM3>  
 F:346-417/Domain: immunoglobulin homology <IMM4>  
 F:436-455/Domain: carboxyl-terminal <CTS>  
 F:14/Disulfide bonds: interchain (to light chain) #status predicted  
 F:28-89/Disulfide bonds: #status experimental  
 F:46-211, 243, 261, 442/Binding site: carbohydrate (asn) (covalent) #status experimental  
 F:136-139, 246-305, 353-415/Dsulfide bonds: #status predicted  
 F:216, 454/Dsulfide bonds: interchain (to heavy chain) #status predicted  
 F:293/Dsulfide bonds: interchain (to mu chain in another subunit) #status predicted

Query Match 10.7%; Score 364; DB 1; Length 455;  
 Best Local Similarity 22.8%; Pred. No. 2.9e-14;  
 Matches 122; Conservative 80; Mismatches 180; Indels 152; Gaps 17;

```

QY 144 ESPPGSSPVQCRSP-RGKNIOGKTLVSQLELDSDSGTWTCTVLQNKKEFKIDIVL 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3 QSPFNVPVLVSCSPSLSDKLVLMGCLARDFLSTISFTN----- 43
QY 203 AFQKASSIVYKKGEOVFSPFLAFTVEKLTSGGELMOMERASSSSKMTFPLDKNEVS 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 44 -----YONNTEVIO-----GIRFPTLRGTGKYLA 68
QY 263 VKRYTQDPKIQM-----GKKLRLHLTPQALPOYAGSGLTLALAKTKGL 308
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 69 TSQVLSPKSLDGSDEYLVCKHYGKNDLHVPIF-----AVAEW 110
QY 309 HOEVLNV-----MRATOLQKNLTCEVWGTPSPKMLSLKLENKEAKVSKREKPVWVLN 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111 NPNVNVVPRPDGSPGAPRKSRLICEATNFTPKRPIVSMKDGVLVSGFTTDPVITEN 170
QY 363 ----PEA-----GMQCLLSDSQVLLBSNIKVLPFTWSTPVEPKSCDK 401
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 171 KSTPTQYKVIISTLTISEIDMLNLYTCRDVHGRGLFLK-----NV 212
QY 402 THTCRPPAPPELLGGPSVFLPFRPKPDLMISRPEVTCVVVDVSHEDPEVKFMWYDGV 461
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 213 SSTCAASSTDL-----FTTIPSPAD-IFLSKSNLTLVSNLATTY- TLNISMASQSG 266
QY 462 EVHNAKTPREBOYNSTYRVVSVLTVLHODMLNGEKYCKVSKNALPAPIEKTISKAKGQ 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 267 EPLETKIKIMESHPIGTFSAKGVASVCVEDMNNRKEFCVYTHHDLRPPQKKFISK----- 322
QY 522 PRE-----POVYTLPPSRDEL-T-KNOVSLTCLVKGFPYSDIAVWESNGQ--PENNYKTT 573
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 323 PNEVNHKHPAYVLLPPAREQNLNRESATVTCLVKGFSPADISVOMLQRGQLLPQEKVYTS 382
QY 574 PVLVD--SDGSFPLYSKLTVDKSRMQGNVSCSWMEALNHNHTOKSLSLSPG 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 383 APMEPRGAPGFYTHSILVTEEBWNSGETYTCVGHGHALPHLVTKTVDKSTG 436

```

RESULT 56  
 A24976  
 Ig mu chain C region (allele b) - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 05-Jun-1988 #sequence\_revision 05-Jun-1988 #text\_change 21-Jan-2000  
 C:Accession: A24976  
 R:Schreier, P.H.; Quester, S.; Bothwell, A.  
 Nucleic Acids Res. 14, 2381-2389, 1986  
 A>Title: Allotypic differences in murine mu-genes.  
 A:Reference number: A24976; MUID:86176735; PMID:3083402  
 A:Accession: A24976  
 A:Molecule type: mRNA  
 A:Residues: 1-455 <SCH>  
 A:Cross-references: GB:X03690; NID:952381; PIDN:CAA27326.1; PID:952382  
 A:Experimental source: strain C57BL/6  
 A>Note: the authors translated the codon AAG for residue 65 as Leu  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 F:346-417/Domain: immunoglobulin homology <IMM>

Query Match 10.7%; Score 364; DB 2; Length 455;  
 Best Local Similarity 22.8%; Pred. No. 2.9e-14;  
 Matches 122; Conservative 80; Mismatches 180; Indels 152; Gaps 17;

```

QY 144 ESPPGSSPVQCRSP-RGKNIOGKTLVSQLELDSDSGTWTCTVLQNKKEFKIDIVL 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3 QSPFNVPVLVSCSPSLSDKLVLMGCLARDFLSTISFTN----- 43
QY 203 AFQKASSIVYKKGEOVFSPFLAFTVEKLTSGGELMOMERASSSSKMTFPLDKNEVS 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 44 -----YONNTEVIO-----GIRFPTLRGTGKYLA 68
QY 263 VKRYTQDPKIQM-----GKKLRLHLTPQALPOYAGSGLTLALAKTKGL 308
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 69 TSQVLSPKSLDGSDEYLVCKHYGKNDLHVPIF-----AVAEW 110
QY 309 HOEVLNV-----MRATOLQKNLTCEVWGTPSPKMLSLKLENKEAKVSKREKPVWVLN 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111 NPNVNVVPRPDGSPGAPRKSRLICEATNFTPKRPIVSMKDGVLVSGFTTDPVITEN 170
QY 363 ----PEA-----GMQCLLSDSQVLLBSNIKVLPFTWSTPVEPKSCDK 401
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 171 KSTPTQYKVIISTLTISEIDMLNLYTCRDVHGRGLFLK-----NV 212
QY 402 THTCRPPAPPELLGGPSVFLPFRPKPDLMISRPEVTCVVVDVSHEDPEVKFMWYDGV 461
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 213 SSTCAASSTDL-----FTTIPSPAD-IFLSKSNLTLVSNLATTY- TLNISMASQSG 266
QY 462 EVHNAKTPREBOYNSTYRVVSVLTVLHODMLNGEKYCKVSKNALPAPIEKTISKAKGQ 521
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 267 EPLETKIKIMESHPIGTFSAKGVASVCVEDMNNRKEFCVYTHHDLRPPQKKFISK----- 322
QY 522 PRE-----POVYTLPPSRDEL-T-KNOVSLTCLVKGFPYSDIAVWESNGQ--PENNYKTT 573
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 323 PNEVNHKHPAYVLLPPAREQNLNRESATVTCLVKGFSPADISVOMLQRGQLLPQEKVYTS 382
QY 574 PVLVD--SDGSFPLYSKLTVDKSRMQGNVSCSWMEALNHNHTOKSLSLSPG 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 383 APMEPRGAPGFYTHSILVTEEBWNSGETYTCVGHGHALPHLVTKTVDKSTG 436

```

RESULT 57  
 EHMS  
 Ig epsilon chain C region (version 1) - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 13-Jun-1983 #sequence\_revision 13-Jun-1983 #text\_change 16-Jul-1999  
 C:Accession: A02144  
 R:Lin, P.T.; Albrandt, K.; Sulejffe, J.G.; Katz, D.H.  
 Proc. Natl. Acad. Sci. U.S.A. 79, 7852-7856, 1982  
 A>Title: Cloning and nucleotide sequence of mouse immunoglobulin epsilon chain cDNA.  
 A:Reference number: A02144; MUID:83117774; PMID:6818553  
 A:Accession: A02144











A:Residues: 320-428 <ZHA>  
A:Cross-references: EMBL:X6593; GB:S38668; NID:g32987  
R:Seno, M.; Kurokawa, T.; Ono, Y.; Ono, H.; Saeeda, R.; Igarashi, K.; Kikuchi, M.; Sugita, T.  
Nucleic Acids Res. 11, 719-726, 1983  
A:Title: Molecular cloning and nucleotide sequencing of human immunoglobulin epsilon chain  
A:Reference number: A93491; MUID:83168897; PMID:6300763  
A:Accession: A93491  
A:Molecule type: mRNA  
A:Residues: 1-428 <SEN>  
A:Cross-references: GB:L00022; GB:U00227; GB:V00555; NID:g185035  
R:Max, E.E.; Batley, J.; Ney, R.; Kirsch, I.R.; Leder, P.  
Cell 29, 691-699, 1982  
A:Title: Duplication and deletion in the human immunoglobulin epsilon genes.  
A:Reference number: A90824; MUID:83001945; PMID:6288268  
A:Accession: A90824  
A:Molecule type: DNA  
A:Residues: 1-358, 'L', 360-428 <MAX>  
A:Cross-references: GB:J00222; NID:g184755  
A:Note: this sequence difference may be due to polymorphism  
R:Bernick, H.H.; Johanson, S.G.O.; Von Behr-Lindstrom, H.  
In Immediate Hyperemstivity: Modern Concepts and Developments, Bach, M.K., ed., pp.1-3.  
A:Reference number: A94418  
A:Accession: A94418  
A:Molecule type: Protein  
A:Residues: 'GAWTL', 6, 'X', 8-16, 'B', 18-43, 'B', 45-52, 55-92, 95-97, 'B', 99-121, 'B', 123, 'L', 124  
A:Experimental source: myeloma protein Nd  
R:Kerten, J.H.; Molgaard, H.V.; Houghton, M.; Derbyshire, R.B.; Viney, J.; Bell, L.O.; G  
Proc. Natl. Acad. Sci. U.S.A. 79, 6661-6665, 1982  
A:Title: Cloning and sequence determination of the gene for the human immunoglobulin epsilon  
A:Reference number: A93933; MUID:83065234; PMID:6815656  
A:Accession: B93933  
A:Molecule type: mRNA  
A:Residues: 1-40; 68-114; 427-428 <KEN>  
A:Cross-references: GB:L00022; NID:g185035  
R:Ikeyama, S.  
FEBS Lett. 224, 306-310, 1987  
A:Title: Purification and characterization of a recombinant human IgE Fc-epsilon fragment  
A:Reference number: S02438; MUID:88083554; PMID:3121387  
A:Accession: S02438  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 98-352 <IKE>  
R:Zhang, K.; Max, E.E.; Cheah, H.K.; Saxon, A.  
J. Biol. Chem. 269, 456-462, 1994  
A:Title: Complex alternative RNA splicing of epsilon-immunoglobulin transcripts produces  
A:Reference number: A53116; MUID:94103254; PMID:8276835  
A:Accession: A53116  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 320-428 <ZHA>  
A:Experimental source: myeloma U266-derived cell line AF-10  
A:Note: sequence extracted from NCBI backbone (NCBI:141701, NCBI:141702)  
R:Hellman, L.  
Eur. J. Immunol. 23, 159-167, 1993  
A:Title: Characterization of four novel epsilon chain mRNA and a comparative analysis of  
A:Reference number: A46536; MUID:93122085; PMID:8439166  
A:Accession: C46536  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 382-426 <HEL>  
A:Cross-references: GB:S55273; NID:g263166; PIDN:AAB24857.1; PID:g263167  
A:Experimental source: B cell myeloma U-266  
A:Note: sequence extracted from NCBI backbone (NCBI:125297)  
A:Accession: D46536  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 382-391 <HE2>  
A:Cross-references: GB:S55276; NID:g263168; PIDN:AAB24858.1; PID:g263169  
A:Experimental source: B cell myeloma U-266  
A:Note: sequence extracted from NCBI backbone (NCBI:125299)  
A:Accession: A46536  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA

A:Residues: 401-428 <HE3>  
A:Cross-references: GB:S53497; MID:g263162; PIDN:AAB24855.1; PID:g263163  
A:Experimental source: B cell myeloma U-266  
A>Note: Sequence extracted from NCBI backbone (NCBIP:123483)  
C:Genetics:  
A:Gene: GDB:IGHB  
A:Cross-references: GDB:119335; OMIM:147180  
A:Map position: 14q32.33-14q32.33  
A:Introns: 1/1; 104/1; 211/1; 319/1  
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into larger superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin F122-87/Domain: Immunoglobulin homology <IM2>  
F128-195/Domain: Immunoglobulin homology <IM2>  
F132-301/Domain: Immunoglobulin homology <IM3>  
F138-407/Domain: Immunoglobulin homology <IM4>  
F141/Dissulfide bonds: Interchain (to light chain) #status predicted  
F145-105,29-85,135-193,239-299,345-405/Dissulfide bonds: #status predicted  
F121,49,99,146,252,275/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F121,209/Dissulfide bonds: Interchain (to heavy chain) #status predicted

Query Match 10.5% Score 359; DB 1; Length 428;  
Best local similarity 27.0%; Pred. No. 5.3e-14;  
Matches 119; Conservative 69; Mismatches 155; Indels 98; Gaps 18;

Dy 267 TQDEPKL-----OMGRKPLRLHLT-----LPQALPOYAGSGNT---TLALEAKTGKL 308  
Db 3 TQSPSVPEPLRCCNKINISNATSVTLGCLATGYFEPEPVMVWDGSLNGTTMTLPATLTLL 62  
Dy 309 ---HOENVLVNRATQLQKKLTCEV-MGPSPKLMLSLKLENKAYSKE---KPYWVL 361  
Db 63 SGHYATISLTLVSGAMAKQMFTCRVAHTPSSTDM----VDNKTFFSCSRDPPTPVTKIL 117  
Dy 362 N-----PEAGMOCCLS-----DSGOVLESNIKYLPTMTSEVPKSC- 399  
Db 118 QSSCDGGGHPRPTIQLCLVSGITPGTIINTWEEDGVMDV----ISTASTQEGELAS 173  
Dy 400 -----DKTHTC-----PPCAPELLGSPSYFLFPKPCKDTLMI 432  
Db 174 TQSELTLQSOKHMLSDRYTCQVTYGCHTFPEDSTRKC-ADSNRPQSAVLSRPSFPD-LFI 231  
Dy 433 SRPEVTCVVVDNSHEDPEVKFNMYVDGEVNAKTKPREEQYNSTYRVVSVLTVLDHQW 492  
Db 232 RKSPTICLVLDLAPSKGTNLTWMSRASGPRVNHSTKEKEQNKGITLVSTLPVGTRDW 291  
Dy 493 LNGEVYCKVSNKALPAPIKITSKAKGQREPOVYTL-----PPSRDELTKNQVSLTCL 547  
Db 292 IEGETVCCRVRTHPHLPALMRSTTKTSGRPAADEVVAFAFTPENPGSDKRT-----LAEL 346  
Dy 548 VKGYEPEDIVAVEMESN--GQPENNYKTPPVLVSDGSFFLYSKLTVDKSMQGNFSSCS 605  
Db 347 IQNMPEPDIGVQMLHNVOLEPARHSTTORPKTKGSGVFVFSLETVRAEMQKDERICR 406  
Dy 606 VMEHALNNHT-QKSLSLSPG 625  
Db 407 AVHEAASPQTQVQRAVSINPG 427

RESULT 63  
S37768  
Ig mu chain C region - human  
C:Species: Homo sapiens (man)  
C:Date: 12-Feb-1998 #sequence\_revision 12-Feb-1998 #next\_change 23-Jul-1999  
C:Accession: S37768  
R:HaIntranactin, N.; Donadel, G.; Sigounas, G.; Notkins, A.L.  
Mol. Immunol. 30, 111-112, 1993  
A>Title: Comparison of complete nucleotide sequence of the human IgM heavy chain constant  
A:Reference number: S37767; MUID:93109369; PMID:8417370  
A:Accession: S37768  
A:Molecule type: mRNA  
A:Residues: 1-453 <HR>  
A:Cross-references: EMBL:X67301; NID:g38407; PIDD:CAA47714.1; PID:g38408







```

Db      61 YAGQIDAMGHGTEIVY---SSASPT-----SPRLYPLSACCSNAPVPAVG 104
      167 KTLVSQV---LEQDSGTWCTVLOQKVERKIDVLVAFQKASIVYKKEGEVSEF 223
Db      105 CLTSPSSAGGSIWEGSG-----GTAAGVSGTGPVKLSF 138
      224 PLAFYVEKLTGSGELWMAERASSSKSWTF-----DLKNKEYSVKRYVQDRLQNGKGL 278
Db      139 VRLSPBEK-----RSPFCSAAPGALAKKEVQYGRV--DP----- 172
      279 PLHLTLPOLPOYAGSGNLTALAEKTKLHGVNLVWNRATQLOKNTLCEVWG--PTSP 336
Db      173 -----VPVAPEVQ-----VLHASSCTPSQSEVELL-----CLVTGSPAPA 210
      337 KMLSLSLKENKAQSKREKPPWVNLPEAGMMOCLISDSGV-----LLESNIKVL--- 387
Db      211 EV-----EWLVGVGGL--LVASQSPAVRSGSTVSLSRVNVSGTD 249
      388 -----PTWSTVPER--KSC-DKTHCTPCPCPAPELLGSPVFLPPKPKDTLM 431
Db      250 WREGKSYSGRVNHPAINTVEVDHVKCCPFGAQSCTSI-----QLYAIIPSPGE-LY 299
      432 ISRTPEVTCVVVDVSHEDPEVKFMYVVDGEVHNNAKTKPR--EEOYNSTYRVSVLTYLH 489
Db      300 ISLDARLRLCLLVNLP-SDSSLSVTMTRE--KSGNLRPDMVQLQEHNGTYSASAVPVST 356
      490 QMNLNKEKYCKVSKNALPAPLEKITSKAKGQPREPQVYTLPPSRDELTKNOVSLTCLVK 549
Db      357 QMNLSEERFTCVQHEHLEPLPSKSYRYRTGPTPLIYFPAHPHEILSLSEVTLSCIVR 416
      550 GFYPSDIAVEM--ESNGQPENNYKTPPVLD-----SDG-SFPLYSKLTVDVDSRMQ 598
Db      417 GRRPRDIERMLRDHAIVATEFTYTAUPEERTANGAGGDDGDTFFVYSKMSVETAKNG 476
      599 GNVFSCSVNHEALHNHYTQKSLSLSPG 625
Db      477 GTVFACMAVHEALPMKRFSGRTLQKQAG 503

RESULT 66
160082
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 23-Jul-1999
C:Accession: I60082
R:Zverev, V.V.; Sidorov, A.V.; Nedospasov, S.A.; Malushova, V.V.; Udalova, I.A.; Andzha
Vopr. Virusol. 40, 100-102, 1995
A:Title: [Nucleotide sequence of two exons of the human T-lymphocyte CD4 receptor gene].
A:Reference number: I60082; MUID:95407135; PMID:7676667
A:Accession: I60082
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-71 <RES>
A:Cross-references: GB:S79267; NID:g1086922; PID:AA835273.1; PID:g1086923
C:Genetics:
A:Intron: 17/1
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology

Query Match      10.5%; Score 357; DB 2; Length 71;
Best Local Similarity 98.6%; Pred. No. 6.9e-15;
Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

      Oy      1 NNRGVPRHLVLVQLALPLPAATQGNKVLGKKGDTVELTCTASQKKSIOFMKNSNQIK 60
      Db      1 NNRGVPRHLVLVQLALPLPAATQGNKVLGKKGDTVELTCTASQKKSIOFMKNSNQIK 60
      Oy      61 ILGNQSGFLTK 71
      Db      61 ILGNQSGFLTK 71

RESULT 67

```

```

EHR1
Ig epsilon chain C region - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change 16-Jul-1999
C:Accession: A93442; A90937; A02143
R:Hellman, L.; Pettersson, U.; Engstrom, A.; Karlsson, T.; Bennich, H.
Nucleic Acids Res. 10, 6041-6049, 1982
A:Title: Structure and evolution of the heavy chain from rat immunoglobulin E.
A:Reference number: A93442; MUID:83064537; PMID:6292865
A:Accession: A93442
A:Molecule type: mRNA
A:Residues: 1-429 <HEL>
A:Experimental source: strain IOU/c/Mel, immunocytoma IR2
R:Kindsvogel, W.R.; Reddy, E.P.; Moore, J.M.; Faust Jr., C.H.
DNA. 1, 335-343, 1982
A:Title: A cloned cDNA probe for rat immunoglobulin epsilon heavy chain: construction, i
A:Reference number: A90937; MUID:83182019; PMID:6820340
A:Contents: myeloma IR162
A:Accession: A90937
A:Molecule type: mRNA
A:Residues: 'N', 169-307, 'L', 309-342 <KIN>
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (Xp
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
F:19-80/Domain: immunoglobulin homology <IM1>
F:118-166/Domain: immunoglobulin homology <IM2>
F:223-291/Domain: immunoglobulin homology <IM3>
F:327-398/Domain: immunoglobulin homology <IM4>
F:46, 99, 170, 240, 265, 369, 419/Binding site: carbohydrate (Asn) (covalent) #statue predicte

Query Match      10.4%; Score 355.5; DB 1; Length 429;
Best Local Similarity 27.8%; Pred. No. 8.5e-14;
Matches 101; Conservative 69; Mismatches 114; Indels 79; Gaps 16;

      Oy      324 KNLTCGEVWGPTSPKLMLSLKENKEAKVSKREKPPVVLNPEAGM----- 367
      Db      74 KNFTCHV--THAPSTFVS-----DLTIARPNINIKPVDLLHSCDNNAFHSTIQ 123
      368 -----WQCLLSD-----SQVLLSNIKVLPTWS--TPVEPKSCKT 402
      Oy      124 YCFVYGHIDNVDSIHW--LMDDRKIYETHAONVLIKEGKLASTSRINITQOQMSRST 181
      Db      403 HTCP-----PCPAPELLGSPVFLPPKPKDTLMISTPVTCTVVVDVSHED 449
      182 FTCKVTSQGENYMAHTRCSDDE--PRGVITLIPSPLD--LYENGTPVLTCLVLDLSEEE 239
      Oy      450 PEVKFNWVVDGEVHNNAKTKPREBOYNSTYRVSVLTVLHODMNLGKCKVSNKALPA 509
      Db      240 -NITWTWRREKKKSGSASQSTKHNATITSLIPDADQWLEGEYQCRVDHPHPK 298
      Oy      510 PIEKTISHAKQPREPQVYTLPPSRDELTKNOVSLTCLVKGFPYSDIAVEM--ESNGQ 566
      Db      299 PIVSITKAPGRKSPAEVYVFLPREEB--KDKRTLCTLIQNFPEDISVQWLQPSKLI 356
      Oy      567 ENNYKTTTPVLDSDGS---FFLYSKLTLYDKSRKQGNVFGSSWNHEALHN-HYQKSL 622
      Db      357 KSQHSHTTP-LKXGNSNORFFIFSRLETKALMTQKTCQFTRVIAHEALREPKLERTISK 415
      Oy      623 SPG 625
      Db      416 SLG 418

RESULT 68
MHUR1
Ig mu heavy chain disease protein (Boc) - human
C:Species: Homo sapiens (man)
C:Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-Jul-1999
C:Accession: A02163
R:Barnikol-Matanabe, S.; Mihaesco, E.; Mihaesco, C.; Barnikol, H.U.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 365, 105-118, 1984
A:Title: The primary structure of mu-chain-disease protein BOR. Peculiar amino-acid sequ

```







C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: alternative splicing; glycoprotein; heterotetramer; immunoglobulin; membrane  
 F:21-90/Domain: immunoglobulin homology <IMM1>  
 F:127-199/Domain: immunoglobulin homology <IMM2>  
 F:237-305/Domain: immunoglobulin homology <IMM3>  
 F:344-415/Domain: immunoglobulin homology <IMM4>  
 F:433-473/Domain: carboxyl-terminal #status predicted <CTS>  
 F:14/Diulfide bonds: interchain (to light chain) #status experimental  
 F:28-88,134-197,244-303,351-413/Diulfide bonds: #status experimental  
 F:46,209,272,279/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:214/Diulfide bonds: interchain (to heavy chain) #status experimental  
 F:291/Diulfide bonds: interchain (to mu chain in another subunit) #status experimental

Query Match 10.3%; Score 352.5; DB 1; Length 473;  
 Best Local Similarity 27.0%; Pred. No. 1,5e-13;  
 Matches 87; Conservative 64; Mismatches 120; Indels 51; Gaps 12;

QY 332 GPTSPKMLSLKLENKAVSKREKPVWLNPEAGMOCILSDSQVLE-SNIKVLPTW 390  
 DB 172 GFTTVKVTSTLTIKESD-----WL---GQSFTCRVDHRLGTFPOGNSSMCVPDQ 218  
 QY 391 STPVEPKSCDKHTCCPCAPPELLGGPSVFLPPLPKDTLMISRTPEVTCVVVDVSHEDP 450  
 DB 219 DTAIR-----VFALPPS-FASIFLTKSTKLCTVLTDLTTYD- 253  
 QY 451 EYKFMNVYDGVENHNAKTRREOVNSTYRVSVLTVLHODMLNGEKYKVSNAKALPAP 510  
 DB 254 SVTISITRONGEAUVKTHNIISSHPNATSAVGEASICEDMNSGRFCTVTHDLPSP 313  
 QY 511 IETKISKAKGP-REFVYTLTPPSRDELTKNOVSLTCLVKGFYPSDIAVEMESNGQP-- 566  
 DB 314 LKQTSIRPGVALHNRDVLRLPAPBQLNRESATITCLVTGFSPADVGVOMQGGQPS 373  
 QY 567 ENNYKTTTPVLD--SDGSFFLYSKLTVDKSRMOQGVFSCVMEHALNHNHTOKSLSP 624  
 DB 374 PEKYVTSAPMPBPQAPGRYFAHSILTVSEEWNTGETYTC-VAHEALPKRVETRVDSK- 431  
 QY 625 GLQDETCAEADGELDGLMTT 646  
 DB 432 ---TEGEVSADEGEENLMT 449

## RESULT 71

MHRB

Ig mu chain C region, secreted form - rabbit  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 13-Aug-1986 #sequence\_revision 13-Aug-1986 #text\_change 23-Aug-1997  
 C:Accession: A02164  
 J:Bernstein, K.E.; Alexander, C.B.; Reddy, E.P.; Mage, R.G.  
 J:Immunol. 132, 490-495, 1984  
 A:Title: Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain of V-Ha2 a  
 A:Reference number: A02164; MUID:84088930; PMID:6418803  
 A:Contents: a2 allotype  
 A:Accession: A02164  
 A:Molecule type: mRNA  
 A:Residues: 1-458 <BR>  
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain diulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin  
 F:21-92/Domain: immunoglobulin homology <IMM1>  
 F:130-203/Domain: immunoglobulin homology <IMM2>  
 F:242-310/Domain: immunoglobulin homology <IMM3>  
 F:349-420/Domain: immunoglobulin homology <IMM4>  
 F:14/Diulfide bonds: interchain (to light chain) #status predicted  
 F:28-90,137-200,249-308,356-418/Diulfide bonds: #status predicted  
 F:46,114,212,261,277,284,445/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:219,457/Diulfide bonds: interchain (to heavy chain) #status predicted  
 F:296/Diulfide bonds: interchain (to mu chain in another subunit) #status predicted

Query Match 10.2%; Score 349; DB 1; Length 458;  
 Best Local Similarity 24.6%; Pred. No. 2.3e-13;  
 Matches 116; Conservative 80; Mismatches 167; Indels 108; Gaps 20;

QY 223 PFL-----AFTVEKLTSGSELMMQABRSASSKMTTDPDKKVEYVKRYTODPKLQWKK 277  
 DB 9 YPLVSCBGLTIDGNLVAMGLARDLFLPSVTFSSW-SFK-KNSELSRVTPTFPVVKRDK 66  
 QY 278 -----LPLHLTLFQAL-----PQYA 292  
 DB 67 YMATQVLVPSKDVLCQTEEVLYCKVQNSNDRDARVFPDSELPRVSVFIPTPDSFS 126  
 QY 293 GSGNITLALQAK-TCKLHQEVNLYVMRATQ-----LQKNTCEVWGFTSPKMLSLKLE 345  
 DB 127 GSGTRKSLRIQATGFSFKQISVSLRBDGQKVESGLTKPVEAEIKGAPATFSSIMLT 186  
 QY 346 NKEAVSRERKPVWLNPEAGMOCILSDSQVLESINIKYLPMTSTVEPKSCDKHTCC 405  
 DB 187 ITESD-----WL---SQSLYTCRVDRHG-IFPDKNVM-----SSQC 219  
 QY 406 PCPAPPELLGGPSVFLPPLPKDTLMISRTPEVTCVVVDVSHEDPEVKFNNYVDGVEVHN 465  
 DB 220 STTPEP-----GIQVPPIAPSPADT-FLSKARLICVLTDLTTYG-SLNISSW-----ASHN 268  
 QY 466 AKT-----KPREOVNSTYRVSVLTVLHODMLNGEKYKVSNAKALPAPLEKTIKAKG 520  
 DB 269 GKALDTNHNITESHNPATFSAVGEASVCAEDMESGEQFTCTVTHADLPFPLKHTISKGR 328  
 QY 521 QPRE-FOYTYTLPPSRDELTKNOVSLTCLVKGFYPSDIAVEMESNGQP--ENNYKTTTPV 576  
 DB 329 VAKHPPAYVLPAPAEQLVRESATITCLVTGFSPADVGVOMQGGQPSADKVTSAAPA 388  
 QY 577 LD--SDGSFFLYSKLTVDKSRMOQGVFSCVMEHALNHNHTOKSLSPG 625  
 DB 389 PEQAPGLYFHSTLTLYTEEDMNSGETTTCVVGHEALPHMTERTVDSKTG 439

## RESULT 72

MHRU

Ig mu chain C region, secreted splice form - human

C:Species: Homo sapiens (man)  
 C:Date: 29-Jul-1981 #sequence\_revision 23-Aug-1997 #text\_change 22-Jun-1999  
 C:Accession: S09357; S16556; S37767; A26243; A26244; I37749; I37750; A02162; B02162  
 R:Dorai, H.; Gillies, S.D.  
 Nucleic Acids Res. 17, 6412, 1989  
 A:Title: The complete nucleotide sequence of a human immunoglobulin genomic C-mu gene.  
 A:Reference number: S09357; MUID:89366890; PMID:2505237  
 A:Accession: S09357  
 A:Molecule type: DNA  
 A:Residues: 1-452 <DOR1>  
 A:Cross-references: EMBL:X14940  
 A:Note: the authors translated the codon AAT for residue 16 as Met  
 R:Dorai, H.  
 submitted to the EMBL Data Library, April 1989  
 A:Reference number: S16556  
 A:Accession: S16556  
 A:Molecule type: DNA  
 A:Residues: 1-39,'L', 41-452 <DOR2>  
 A:Cross-references: EMBL:X14940  
 R:Hatindranath, N.; Donadel, G.; Sigounas, G.; Nockins, A.L.  
 Mol. Immunol. 30, 111-112, 1993  
 A:Title: Comparison of complete nucleotide sequence of the human IgM heavy chain constan  
 A:Reference number: S37767; MUID:93109369; PMID:841370  
 A:Accession: S37767  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-190,'S',192-414,'V',415-452 <HAR>  
 A:Cross-references: EMBL:X67292; NID:G38405; PIDN:CAA47708.1; PID:G38406  
 R:Rabbits, T.H.; Forster, A.; Milstein, C.P.  
 Nucleic Acids Res. 9, 4509-4524, 1981  
 A:Title: Human immunoglobulin heavy chain genes: evolutionary comparisons of C-mu, C-delta  
 A:Reference number: A26243; MUID:82059479; PMID:6795593  
 A:Accession: A26243  
 A:Molecule type: DNA  
 A:Residues: 1-17,105-186,200-259,296-322,339-416,'D',418-452 <RAB>  
 A:Cross-references: GB:X01310; NID:G184715



R:Doibly, T.W., Devono, J., Croce, C.M.,  
Proc. Natl. Acad. Sci. U.S.A. 77, 6027-6031, 1980  
A:Title: Cloning and partial nucleotide sequence of human immunoglobulin mu chain cDNA  
A:Reference number: A26244; MUID:81077306; PMID:6777778  
A:Accession: A26244  
A:Molecule type: mRNA  
A:Residues: 298-366/436-452 <DOL>  
A:Cross-references: GB:J00257; NID:g185053; PIDN:AA53508.1; PID:g185056; GB:J00258; NID:J00258  
R:Takehashi, N., Nakai, S., Honjo, T.  
Nucleic Acids Res. 8, 5963-5991, 1980  
A:Title: Cloning of human immunoglobulin mu gene and comparison with mouse mu gene.  
A:Reference number: 137748; MUID:81124312; PMID:6450943  
A:Accession: 137749  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 347-370 <TRK1>  
A:Cross-references: EMBL:V00562; NID:g33448; PIDN:CAA23825.1; PID:g929649  
A:Accession: 137750  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 433-452 <TRK2>  
A:Cross-references: EMBL:V00563; NID:g33454; PIDN:CAA23826.1; PID:g825684  
R:Milnesco, E., Barnikol-Watanabe, S., Barnikol, H.U., Milnesco, C., Hilschmann, N.  
Eur. J. Biochem. 111, 275-286, 1980  
A:Title: The primary structure of the constant part of mu-chain-disease protein BOT.  
A:Reference number: A02162; MUID:81066716; PMID:6777162  
A:Contents: Mu-chain-disease protein Bot and revisions to sequence of Gal  
A:Accession: A02162  
A:Molecule type: protein  
A:Residues: 1-17, 'BPS', '22', 'T', '24-82', 'N', '84-90', 'Z', '92-93', 'B', '95', 'B', '97-144', 'E', '146-162', 'E', '164-165', 'E', '166-214', 'G', '216-262', 'D', '264-295', 'D', '297-414', 'A', '415-416', 'A', '417-418', 'A', '419-420', 'A', '421-422', 'A', '423-424', 'A', '425-426', 'A', '427-428', 'A', '429-430', 'A', '431-432', 'A', '433-434', 'A', '435-436', 'A', '437-438', 'A', '439-440', 'A', '441-442', 'A', '443-444', 'A', '445-446', 'A', '447-448', 'A', '449-450', 'A', '451-452', 'A', '453-454', 'A', '455-456', 'A', '457-458', 'A', '459-460', 'A', '461-462', 'A', '463-464', 'A', '465-466', 'A', '467-468', 'A', '469-470', 'A', '471-472', 'A', '473-474', 'A', '475-476', 'A', '477-478', 'A', '479-480', 'A', '481-482', 'A', '483-484', 'A', '485-486', 'A', '487-488', 'A', '489-490', 'A', '491-492', 'A', '493-494', 'A', '495-496', 'A', '497-498', 'A', '499-500', 'A', '501-502', 'A', '503-504', 'A', '505-506', 'A', '507-508', 'A', '509-510', 'A', '511-512', 'A', '513-514', 'A', '515-516', 'A', '517-518', 'A', '519-520', 'A', '521-522', 'A', '523-524', 'A', '525-526', 'A', '527-528', 'A', '529-530', 'A', '531-532', 'A', '533-534', 'A', '535-536', 'A', '537-538', 'A', '539-540', 'A', '541-542', 'A', '543-544', 'A', '545-546', 'A', '547-548', 'A', '549-550', 'A', '551-552', 'A', '553-554', 'A', '555-556', 'A', '557-558', 'A', '559-560', 'A', '561-562', 'A', '563-564', 'A', '565-566', 'A', '567-568', 'A', '569-570', 'A', '571-572', 'A', '573-574', 'A', '575-576', 'A', '577-578', 'A', '579-580', 'A', '581-582', 'A', '583-584', 'A', '585-586', 'A', '587-588', 'A', '589-590', 'A', '591-592', 'A', '593-594', 'A', '595-596', 'A', '597-598', 'A', '599-600', 'A', '601-602', 'A', '603-604', 'A', '605-606', 'A', '607-608', 'A', '609-610', 'A', '611-612', 'A', '613-614', 'A', '615-616', 'A', '617-618', 'A', '619-620', 'A', '621-622', 'A', '623-624', 'A', '625-626', 'A', '627-628', 'A', '629-630', 'A', '631-632', 'A', '633-634', 'A', '635-636', 'A', '637-638', 'A', '639-640', 'A', '641-642', 'A', '643-644', 'A', '645-646', 'A', '647-648', 'A', '649-650', 'A', '651-652', 'A', '653-654', 'A', '655-656', 'A', '657-658', 'A', '659-660', 'A', '661-662', 'A', '663-664', 'A', '665-666', 'A', '667-668', 'A', '669-670', 'A', '671-672', 'A', '673-674', 'A', '675-676', 'A', '677-678', 'A', '679-680', 'A', '681-682', 'A', '683-684', 'A', '685-686', 'A', '687-688', 'A', '689-690', 'A', '691-692', 'A', '693-694', 'A', '695-696', 'A', '697-698', 'A', '699-700', 'A', '701-702', 'A', '703-704', 'A', '705-706', 'A', '707-708', 'A', '709-710', 'A', '711-712', 'A', '713-714', 'A', '715-716', 'A', '717-718', 'A', '719-720', 'A', '721-722', 'A', '723-724', 'A', '725-726', 'A', '727-728', 'A', '729-730', 'A', '731-732', 'A', '733-734', 'A', '735-736', 'A', '737-738', 'A', '739-740', 'A', '741-742', 'A', '743-744', 'A', '745-746', 'A', '747-748', 'A', '749-750', 'A', '751-752', 'A', '753-754', 'A', '755-756', 'A', '757-758', 'A', '759-760', 'A', '761-762', 'A', '763-764', 'A', '765-766', 'A', '767-768', 'A', '769-770', 'A', '771-772', 'A', '773-774', 'A', '775-776', 'A', '777-778', 'A', '779-780', 'A', '781-782', 'A', '783-784', 'A', '785-786', 'A', '787-788', 'A', '789-790', 'A', '791-792', 'A', '793-794', 'A', '795-796', 'A', '797-798', 'A', '799-800', 'A', '801-802', 'A', '803-804', 'A', '805-806', 'A', '807-808', 'A', '809-810', 'A', '811-812', 'A', '813-814', 'A', '815-816', 'A', '817-818', 'A', '819-820', 'A', '821-822', 'A', '823-824', 'A', '825-826', 'A', '827-828', 'A', '829-830', 'A', '831-832', 'A', '833-834', 'A', '835-836', 'A', '837-838', 'A', '839-840', 'A', '841-842', 'A', '843-844', 'A', '845-846', 'A', '847-848', 'A', '849-850', 'A', '851-852', 'A', '853-854', 'A', '855-856', 'A', '857-858', 'A', '859-860', 'A', '861-862', 'A', '863-864', 'A', '865-866', 'A', '867-868', 'A', '869-870', 'A', '871-872', 'A', '873-874', 'A', '875-876', 'A', '877-878', 'A', '879-880', 'A', '881-882', 'A', '883-884', 'A', '885-886', 'A', '887-888', 'A', '889-890', 'A', '891-892', 'A', '893-894', 'A', '895-896', 'A', '897-898', 'A', '899-900', 'A', '901-902', 'A', '903-904', 'A', '905-906', 'A', '907-908', 'A', '909-910', 'A', '911-912', 'A', '913-914', 'A', '915-916', 'A', '917-918', 'A', '919-920', 'A', '921-922', 'A', '923-924', 'A', '925-926', 'A', '927-928', 'A', '929-930', 'A', '931-932', 'A', '933-934', 'A', '935-936', 'A', '937-938', 'A', '939-940', 'A', '941-942', 'A', '943-944', 'A', '945-946', 'A', '947-948', 'A', '949-950', 'A', '951-952', 'A', '9

QY	332	GPTSPKMLSLKLENKKAQVSRKREKPVNLPNPAQMOCLLSDSGQVILE-SNFKVLPTW	390
Db	172	GPTTKKVVSTLTIKKSD-----WL-----GQSFCTGVDRGLTFQOQASMCVPDQ	218
QY	391	STPVEPKSCDKTKHTCPPCPAPELLGSPSVFLPFPKPKDITLMISRTPEVTCVVDVSHEDP	450
Db	219	DTAIR-----VFAPPS-FASIFLTKSTKLTLVTLDTTYD-	253
QY	451	EKFEMWYDGVGVHNAKTPRREQNSTRVSVLVTLHODVNLGKRYKCVSKALPAP	510
Db	254	SVTISMTQNGAVKTHNTNIISSHPNATFSVAGEASICEDDNMSGKGFCTVTHTDLPSP	313
QY	511	IEKTSIKAKGCP-REPOVYTLTPPSRDELTP-KNOVSTLCVKGFPSPDIAYEASNGCP--	566
Db	314	LKQTSIRPRGVALNHRPDVLLPAPRQGLNLRBATITTCVLTGSPAPVPMQMRGQPLS	373
QY	567	ENNYKTPPEVLD-SDGSFFLYSKLTVDKSRMOQGVFSCSVNHEALNHYTKSLSISP	624
Db	374	PEKYTASAPMEPQAPGRYFAHSILTVSEEWMTGETYTC-VAHEALPNRTERTVDKST	432
QY	625	G 625	
Db	433	G 433	

RESULT 73

S03961

Ig mu chain C region - house shrew (fragment)

C:Species: Suncus murinus (house shrew)

C:Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 21-Jan-2000

C:Accession: S03961

R:Shiguro, H.; Ichihara, Y.; Namikawa, T.; Nagatsu, T.; Kurosawa, Y.

FEBS Lett. 247, 317-322, 1989

A:Title: Nucleotide sequence of Suncus murinus immunoglobulin mu gene and comparison with

A:Reference number: S03961; M0ID:89232144; PMID:2497033

A:Accession: S03961

A:Molecule type: DNA

A:Residues: 1-457 <ISH>

A:Cross-references: EMBL:X13920

C:Genetics:

A:Introns: 106/1; 221/1; 327/1

C:Superfamily: immunoglobulin C region, immunoglobulin homology

C:Keywords: immunoglobulin

F,241-309/Domain: immunoglobulin homology <IMM>

Query Match	10.2%;	Score 348;	DB 2;	Length 457;
Best Local Similarity	23.7%;	Pred. No. 2.6e-13;		
Matches 109;	Conservative 90;	Mismatches 160;	Indels 100;	Gaps 18;

QY	204	FOKASSIVYKKEGEQVFSFPLAFVTEKLTGSGELMWQABR-ASSSKSWITFDLKN----	258
Db	43	FNSSSI-----SSQNIYVFPVFTGCKYMATSGVLLPSTAILQSTDDYITCHKHTTGE	97
QY	259	KEVSVKRVTDQPKLQMGKLLPLHLT--POLPOVAGSGMLTALBAKTKGLHGEVNLV	316
Db	98	KEKKY-----ELQYTPELPPNVSIFFP---PRNSFSGN-----HPTSQLI	135
QY	317	MRATLOLQNLTECV-----GPTSPKMLSLKLENKKAQVS	352
Db	136	COASGSPSPRTIVMSLQRCGEVPQPSLVSTGSAVEAPKSGSPPTTFVRSRLTTENE----	191
QY	353	KREKPVWVLPNPAQMOCLLSDSGQVLLSNIKVLPTWSTPVEPKSCDKTKHTCPPCPAPE	412
Db	192	-----WLSQRE-----FTQALHKG-LTFQKNVSV-----CMGDD	221
QY	413	LLGSPSVFLPFPKPKDITLMISRTPEVTCVVDVSHEDPEVYKFWYDGVGVHNAKTPRE	472
Db	222	TSTGISVFLPPTFAN-IFLTQSAQLTCLVTGATYD-SLDISRSRNGEALQTHVNISE	279
QY	473	EOYNSITVAVSVLTVLHODVNLGKRYKCVSKALPAPIKTSIKAKGQPRE-QQVYTLF	531
Db	280	SHPNSTFAKGNASVCREWESGEKFTCVQSHDLSPLKQSLSRPVDVANDPSPVFLP	339











```

RESULT 79
S60266
novel antigen receptor precursor - nurse shark
C:Species: Ginglymormora cirratum (nurse shark)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 21-Jul-2000
C:Accession: S60266
R:Greenberg, A.S.; Avila, D.; Hughes, M.; Hughes, A.; McKlinney, E.C.; Flajnik, M.F.
Nature 374, 168-173, 1995
A:Title: A new antigen receptor gene family that undergoes rearrangement and extensive
A:Reference number: S60266; MID:95183140; PMID:7877689
A:Accession: S60266
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-684 <GRE>
A:Cross-references: EMBL:U8701; MID:g659442; PID:AA848195.1; PID:g659443

Query Match
Best Local Similarity 23.5%; Score 330.5; DB 2; Length 684;
Matches 158; Conservative 83; Mismatches 237; Indels 195; Gaps 29;

QY 16 LALLPAAQGNKVVLOKGDVTELTCTAS--QKKSIFHW-KNSNLIKLGSGSFLTKG 72
DB 145 VLLSHATEQRA-----NRVQLVCLISGYYPENIAVSMQNTKTI-----TSGFATTS 194
QY 73 PSKLNDRADSRSL-----WDQGNFPLIKLIKIEDSDTYICEVEDQKEEYQLLVFGL 125
DB 195 PVKTSNDSSCASLKVPIQEMSRGS-----VYSCVY----- 226
QY 126 TANSDTLLQGSGLTLLTSPSSPSVOCSPRGNIGGKTLVSQLELODSTGTCT 185
DB 227 -SHSATSSNQKREIRST-----SEIAVLLRD-----TVEEIMDKSATLVCE 268
QY 186 VIQONKKEVFKIDIVLAQKASSIYKKEGEQVE--FSPLAFVKEKLTSGSELIMQA 242
DB 269 VISTVSAG-----VVVSMWNGKV--RREGVQMETKMSGNQYLISLTSVVEW--- 317
QY 243 ERASSKSWITFDLKNKEVSVKRVTDOPKLOMGKKLPLHLTLPOALPOYAGSGNLTAL 302
DB 318 ---QSGVEYTCSAKODQSTPVRVKTTRKARVEFTPR-HLRLLPSPSEIQTST----- 366
QY 303 AKTGKLHGEVNLVVMKATQLQKNLTCEVWGPTSPKLMLSLKLENKAKVSKREKPYVVLN 362
DB 367 -----SATLTCLIRGFYDPKVS----- 385
QY 363 PRAGMWQCLLSDGVLLESNIKVLPTM-----STPEPKSCCKTHTC--- 405
DB 386 -----WQ-----KDDVSVSANVTNPTALEODLTSTSLMLVLAEMKSGAK-YTCTAS 434
QY 406 -PP-----CPAPELLGSPVFLPPPKKDTLMISRTPEVTCVVVDVSHED 449
DB 435 HPSQSSTVKRVIRNKNVDCRQTDI---SVSLKRP-PFEIWTQQTATTICEIV---YSD 486
QY 450 PE-VKFNMYVDGVEVHNAKTPREQYNSTYRVSVLTVLHODWLNKGEYKCKVSNKALP 508
DB 487 LENIKVFQVNGVERKKGVEQTONPEWSSKSTIVSLKXMASBEMDSGTETVCLVEDSELP 546
QY 509 AIEKTIKSAK-GQREPOVYTLPSRDEL-TKNQVSLCTLYKGYPPSIAVEMBSNGO- 565
DB 547 TVPKASIRKANYSONMPKPYVLLHPSTDIIDENSATLMCLNTNFPBAIYVGMAMNDLT 606
QY 566 PENNYKTPPVLDSDGSPFLYSKLTVDKSRMOQGNVFCSVNHEALH-----NHYYQK 618
DB 607 LDGSRITQVDSKSGSSFTVRLRLTAEMNSDTTYSCLVGHPSLNDRLINSTKSNCK 666
QY 619 SLSLSPGLQDDET 631
DB 667 PPSVNVSVVLSDT 679

```

```

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Sep-1999
C:Accession: S31436
R:Fellah, J.S.; Miles, M.V.; Schwager, J.; Charlemagne, J.
Submitted to the EMBL Data Library, November 1992
A:Description: cDNA sequence of Ambystoma mexicanum upslon heavy Igy chain.
A:Reference number: S31436
A:Accession: S31436
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-433 <FEU>
A:Cross-references: EMBL:X69492; MID:G62420; PID:G62421
C:Superfamily: immunoglobulin C region; immunoglobulin homology

Query Match
Best Local Similarity 23.6%; Score 313.5; DB 2; Length 433;
Matches 106; Conservative 64; Mismatches 162; Indels 117; Gaps 19;

QY 216 GEVFEFPPLAFTVEKLTGSGELMWQARASSKSS-----WTF-----FDLKNE 260
DB 34 GEVFEVSKAGSTT---MPGKTFPAQQAQTALSTSSQIRIPAYEMKTNSSVSCVKKRP 90
QY 261 VSV-----KRVTDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLH 309
DB 91 TSEIHKTTTSAECKKATSKESVQV-----LQSCADPTDNGSI----- 129
QY 310 QEVNLVVMKATQLQKNLTCEVWGPTSPKLMLSLKLENKAKVSKREKPYVVLNPRAGMW 369
DB 130 -----ELVCLISGYTPDNIQVRWLVDNKAAPLQGTSP-----PKG--- 165
QY 370 CLLSDGQVLLESNIKVLPT-WST-----PVPKSCDKHTHTCPCPAPILGAP 417
DB 166 ---DGQGFPTTSQINVTKSDMASGDKYTCVYEHATSSRAEDITHNCADQTFP---QP 219
QY 418 SVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEYKFNWYVDGVEVHNAKTPRE----- 472
DB 220 KVFLLAPKARD-LYIANQPVVYLCITKMNEND-SLSVTW-----KREGPEBAVYS 269
QY 473 EGY---NSTYRVSVLTVLHODWLNKGEYKCKVSNKALPAPIEKTIKSAKQRPPOVYT 529
DB 270 EGYIDSDGTFPAMSLNLTNNKEMERGDFTCKVKHFDLPFLPSRSVSKPTGSAFTWYV 329
QY 530 LPPSRDELTK-NOVELTCLVKGFPSPDIJAVEMBSNGO--PENNYKTPPVLDSD-----GS 582
DB 330 FAPHEMELANDYFVSULTLVKVSFSPDIIYQMKQKSVYPSDKYVSMRQEAAGTAGCT 389
QY 583 FFLYSKLTVDKSRMOQGNVFCSVNHEAL 611
DB 390 YFSYSMLTIQKSDMDKRETFCTVAHSAV 418

RESULT 81
S21461
T-cell surface glycoprotein CD4 (allele 1) - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 21-Jan-2000
C:Accession: I47131; S21461
R:Gueneatson, K.; Germana, S.; Sundt, T.M.
J. Immunol. 151, 1365-1370, 1993
A:Title: Extensive allelic polymorphism in the CDR2-like region of the miniature swine
A:Reference number: I47131; MID:93329116; PMID:8335933
A:Accession: I47131
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-99 <GU2>
A:Cross-references: EMBL:X65629; MID:G1928; PID:CAA46583.1; PID:G388232
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C:Keywords: glycoprotein; T-cell
F:3-81/Domain: immunoglobulin homology <IMM>

Query Match
Best Local Similarity 60.2%; Score 305.5; DB 2; Length 99;
Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

```



Oy 32 KKGDTELTCTAASQKSIQFMKNSNQIKLGNQSF--TKGPKSLNDRADSRSLMDQG 90  
 Db 1 KAGDLAEIPCHSSQKKNLPFWNKNSNQTKILGHSFWHTASVTELTSLDSKKNMWDG 60  
 Oy 91 NEPLIKKLIKEDSDTYICEVEDQKEEYOLVFGLTAN 128  
 Db 61 SFLIKKLEVDTSIGYICEVEDKRIEYOLVFLTLAS 98

## RESULT 82

AHRB

Ig alpha chain C region - rabbit (fragment)  
 C/Species: Oryctolagus cuniculus (domestic rabbit)  
 C/Date: 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 16-Jul-1999  
 C/Accession: A02174  
 R/Knight, K.L.; Martens, C.L.; Stoklosa, C.M.; Schneiderman, R.D.  
 Nucleic Acids Res. 12, 1657-1670, 1984  
 A/Title: Genes encoding alpha-heavy chains of rabbit IgA: characterization of cDNA encod  
 A/Reference number: A02174; MUID:84144059; PMID:6322114  
 A/Accession: A02174  
 A/Molecule type: mRNA  
 A/Residues: 1-299 <KNT>  
 A/Cross-references: GB:X00351; NID:g1575; PIDN:CAA25100.1; PID:g1576  
 C/Comment: This immunoglobulin belongs to the IgA-g subclass. It was isolated from a rat  
 C/Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa  
 chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 16  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C/Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin; plasma  
 F:189-261/Domain: immunoglobulin homology <IM2>  
 F:138-286/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.9%; Score 305; DB 1; Length 299;  
 Best Local Similarity 31.6%; Pred. No. 5.2e-11;  
 Matches 85; Conservative 44; Mismatches 120; Indels 20; Gaps 11;  
 Oy 369 QCLSDSGQVLLSENIKVLPTWSTPVEPKSCDKHTPCPCPAP--ELLG--PSVFLFP 423  
 Db 20 QCLGKSAACHVEYNSVI--NESLPVPDPCCPANSCTCPSSSRNLSGCGSLSLQ 77  
 Oy 424 PKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKRREQDYNSTYRV 483  
 Db 78 PDLG--LILGRDASLTCTLSGLKNPEDAV--FTW--EPTNGNEPVQQAQRLDSCYSVS 133  
 Oy 484 VLTVLHODMLNGEKYCKVSNKALPAP-PIETISKAKQPREPOVYTLPSRDELTKN-Q 541  
 Db 134 VLPSSAETWKAARTEFTCTVHPEIDSGSLTATISRGVTP--POVHLPPSEBELALNEQ 191  
 Oy 542 VSLTCLVKGFPSPDIAYVESNGQ--PENNY--KTPPVLDSDGSFFLYSKLTVYDKSRW 596  
 Db 192 VTLTCLVGRGSPKDVLSWMHQGEVEDSFLVWKMSPSSQDKATYATISLILVPAEDW 251  
 Oy 597 QCGNVFSCVMHEALHNHYTOKSLSLSPG 625  
 Db 252 NQGDYSCVMGHGLAEHFTORTIDRLAG 280

## RESULT 83

S09269

Ig alpha chain C region - rabbit (fragment)  
 C/Species: Oryctolagus cuniculus (domestic rabbit)  
 C/Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 16-Jul-1999  
 C/Accession: S09269  
 R/Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.  
 EMBO J. 8, 4041-4047, 1989  
 A/Title: The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13  
 A/Reference number: S09264; MUID:90076124; PMID:5512120  
 A/Accession: S09269  
 A/Status: not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-357 <BUR>  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C/Keywords: immunoglobulin

F:142-208/Domain: immunoglobulin homology &lt;IM4&gt;

Query Match 8.9%; Score 304.5; DB 2; Length 357;  
 Best Local Similarity 33.3%; Pred. No. 7e-11;  
 Matches 93; Conservative 39; Mismatches 116; Indels 31; Gaps 15;  
 Oy 363 PEAGMOCCLIS---DSGQ-VLSENIKVLPTWSTPVEPKSCDKHTCP-PCPABELLGGP 417  
 Db 75 PEENSVAACHVEHNHYDKQGHVTPS----PPECOPPTPCPSDTP--TCCPCPSPS-CGEP 127  
 Oy 418 SVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKRREQDYN 477  
 Db 128 SLSLQRFPLPD-LILNSNASLTCTLSGLKNPEDAV--FTW----EPTNG-NKPVQGSVQS 179  
 Oy 478 ----TYRVSVLTVLHODMLNGEKYCKVSNKALP-APIETISKAKQPREPOVYTLPP 532  
 Db 180 YPGCGYSVSVLPCCAPPMNAGTEFTCTVHPEIDGFLTRAKISKDQGAIIIPGVHLDP 239  
 Oy 533 SRDELTKNQ-VSLTCLVKGFPSPDIAYVESNGQ--QENNY--KTPPVLDSDGSFFLY 586  
 Db 240 PSEELALNELVTLTCLVGRGSPKDVLYWTKNGVNVENSFLVWKPLPBPQGEPTTAAVT 299  
 Oy 587 SKLTVDSRWQCGNVFSCVMHEALHNHYTOKSLSLSPG 625  
 Db 300 SLKVPADWNNQNESYTCVVGHGLAEHFTORTIDRLSG 338

## RESULT 84

S09276

Ig alpha chain C region - rabbit (fragment)  
 C/Species: Oryctolagus cuniculus (domestic rabbit)  
 C/Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 16-Jul-1999  
 C/Accession: S09276  
 R/Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.  
 EMBO J. 8, 4041-4047, 1989  
 A/Title: The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13  
 A/Reference number: S09264; MUID:90076124; PMID:5512120  
 A/Accession: S09276  
 A/Status: not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-338 <BUR>  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C/Keywords: immunoglobulin  
 F:228-300/Domain: immunoglobulin homology <IM4>

Query Match 8.8%; Score 302; DB 2; Length 338;  
 Best Local Similarity 31.2%; Pred. No. 9.2e-11;  
 Matches 84; Conservative 45; Mismatches 120; Indels 20; Gaps 11;  
 Oy 369 QCLSDSGQVLLSENIKVLPTWSTPVEPKSCDKHTPCPCPAP--ELLG--PSVFLFP 423  
 Db 59 QCLGKSAACHVEYNSVI--NESLPVPDPCCPANSCTCPSSSRNLSGCGSLSLQ 116  
 Oy 424 PKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKRREQDYNSTYRV 483  
 Db 117 PDLG--LILGRDASLTCTLSGLKNPEDAV--FTW--EPTNGNEPVQQAQRLDSCYSVS 172  
 Oy 484 VLTVLHODMLNGEKYCKVSNKALPAP-PIETISKAKQPREPOVYTLPSRDELTKN-Q 541  
 Db 173 VLPSSAETWKAARTEFTCTVHPEIDSGSLTATISRGVTP--POVHLPPSEBELALNEQ 230  
 Oy 542 VSLTCLVKGFPSPDIAYVESNGQ--PENNY--KTPPVLDSDGSFFLYSKLTVYDKSRW 596  
 Db 231 VTLTCLVGRGSPKDVLSWMHQGEVEDSFLVWKMSPSSQDKATYATISLILVPAEDW 290  
 Oy 597 QCGNVFSCVMHEALHNHYTOKSLSLSPG 625  
 Db 291 NQGDYSCVMGHGLAEHFTORTIDRLAG 319

## RESULT 85

HVRC2

Ig mu chain C region (clone 12022) - horn shark (fragment)







Ig alpha chain - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 18-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 26-Aug-1999  
C:Accession: A46507  
R:Mansikka, A.  
J:Immunol. 149, 855-861, 1992  
A>Title: Chicken IgA H chains. Implications concerning the evolution of H chain genes.  
A:Reference number: A46507; MUID:92340889; PMID:1634774  
A:Accession: A46507  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-585 <MAN>  
A:Cross-references: GB:S40610; NID:9251907; PID:9251908  
A>Note: sequence inconsistent with the nucleotide translation  
A>Note: sequence extracted from NCBI backbone (NCBIN:109906, NCBIPI:109907)  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
F:34-117/Domain: immunoglobulin homology <IMM>

Query Match 8.7%; Score 295.5; DB 2; Length 585;  
Best Local Similarity 24.3%; Pred. No. 4.5e-10;  
Matches 161; Conservative 87; Mismatches 264; Indels 151; Gaps 35;

```
QY 10 LLLVQLALLPRAAGNVLGKK--GDTVELTCTAS--QKKSIOFHH-----KNSNOI 59
Db 8 LLLLAALPGLMAAVTLTDSGGGLQTPGALSLVCKASGFTSSVGMWRQAPGKLEWV 67
QY 60 KILNQSGFLTKPSPKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQ 119
Db 68 AGIDDDSGTGCGVAVGQARATISRDN--GQSTVRLQLNMLAEDATATYCAKG----- 118
QY 120 LLVRLGTLNSDTHLLIQGSLTLTLESPPGSSPSV-----QCRSPRKNIOGKTLISV 171
Db 119 --ASGCGACASIDAMGHTGTEIVSSASASRPTLQLPLPSDCDP---VNTIG-CLVT 172
QY 172 SQLELDGSGTCTYLOKQKVEFKIDIVLAFOKASSIVYKKEBQVBFSPPLATYRK 231
Db 173 SFLPPVTVTTGGAADATV-----TSLPVATTGGIVSLTALTIVPEEQ 218
QY 232 LTGSGELMWOAERASSSKSWITFDLKNK-----EVSVKRVT--QDPKIQMGKLLPLHL 282
Db 219 LQGN-EFVCRAGHATGA-----DYKETIGDGVCPIRTSKVTLISDPTQDFFERRVLLV 272
QY 283 TLPOALPOYAGSGNLTALAEKTKLHQNVLVNRATQLOKNLTCEVWGPTSPKMLSL 342
Db 273 CLVEGLPS-AGA-----AIQ-----W----- 287
QY 343 KLEKKEAVSKREKRVWVLNPEAGMOCLLSDSGOVLLESNIKVL-PTWSTPVEPKSCDK 401
Db 288 -LQDNEKTPAPE-----SDSGGSDC--TESG-VTQMSRVNVTYRKSMEGAGQ-FGCRV 336
QY 402 TH-----TCPPCPA-PELLGPGSVFLPPKPKDTLMISRTPEVTCVAVVDVSHEDP 450
Db 337 THGALKEPVTAIVTSDCATPOL-----QVSLPLPLEE-LVSNHATVTCVSNAAAAD- 390
QY 451 EVKFNWYVD---GVEVNAKTKPREEQYNSTYRVSVLTVLHQMNLNGEKYKCVSNKAL 507
Db 391 GVSWSMSRSSGGGLDV-----SQTEDRQADGRYTVASFIRVCAEEMNGETGCGSVREGV 446
QY 508 PAPLEKTSKAKGP-REPOVYTLPSRDELTL-KNOVSLTCLVKGFTYSDIIVAVMESNGQ 565
Db 447 -VVAESIRKETDTPPLHAPSVYVPPPAEELSLQETATLTMASSFLPSSILLTWQOQ 505
QY 566 P-ENNYKTPPVLDSDGSFF-LYSKLTVDKSRMQQGNVFCSVMEALNHNHYTKQKSL 622
Db 506 PISQNYILIFGP--EKQDGFYSLYKLVSVEDMQRGVFGCVGHDOIPLNFIHKSIDK 563
QY 623 SPG 625
Db 564 NAG 566
```

RESULT 89  
HVRCO

Ig mu chain C region, membrane-bound (clone 3050) - horn shark  
C:Species: Heterodontus francisci (horn shark)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Aug-1996  
C:Accession: S01854; C32716; A66530  
R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloett, M.J.; Litman, G.W.  
EMBO J. 7, 1979-1988, 1988  
A>Title: Complete structure and organization of immunoglobulin heavy chain constant region  
A:Reference number: S00980; MUID:88328985; PMID:3138109  
A:Accession: S01854  
A:Molecule type: DNA  
A:Residues: 1-461 <KOK>  
A:Cross-references: EMBL:X07781  
A>Note: the sequence was determined from the germ-line gene  
R:Kokubu, F.; Hinds, K.; Litman, R.; Shambloett, M.J.; Litman, G.W.  
Proc. Natl. Acad. Sci. U.S.A. 84, 5868-5872, 1987  
A>Title: Extensive families of constant region genes in a phylogenetically primitive vert  
A:Reference number: A32716; MUID:87289703; PMID:3475706  
A:Accession: C32716  
A:Molecule type: DNA  
A:Residues: 1-99 <KO2>  
A:Cross-references: GB:M17186  
C:Genetics:

A:Introns: 100/1, 206/1, 309/1, 419/1, 459/3  
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin  
F:123-190/Domain: immunoglobulin homology <IMM2>  
F:228-231/Domain: immunoglobulin homology <IMM3>  
F:330-400/Domain: immunoglobulin homology <IMM4>  
F:438-458/Domain: transmembrane #status predicted <TM>  
F:27-85,130-188,235-289,337-398/Disulfide bonds: #status predicted  
F:164,200,245,275,374,411,415,437/Binding site: carbohydrate (Aan) (covalent) #status pr

Query Match 8.6%; Score 295; DB 1; Length 461;  
Best Local Similarity 25.5%; Pred. No. 3.6e-10;  
Matches 121; Conservative 69; Mismatches 167; Indels 118; Gaps 24;

```
QY 208 SSIYKKEGQVER--SFLPAF-----VEKLTSGLMWOAERASSKSW 251
Db 37 TSVSKQDNBPITGTLKTPSVLNKKGTYYQSSQLTTEBVSCKTYCERRGES--VW 94
QY 252 ITFDLKNKEVSVKRVNTPDPLQMGKLLPLHLTLPOALPOYAGSGNLTALAE----- 303
Db 95 I-----KEI-----PDCK---GDKNHPTVILLQSSSEITSRFATVLCSDIFPES 139
QY 304 -----KTGKLHQNVLVNRATQLOKNLTCEVWGPTSPKMLSLKLEKKEAVSKREK 357
Db 140 ITVSWLKDGO-HMESGFV-----SPTCGVNGTFSATSRLT-----VPAE-- 179
QY 358 VWVLNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVEPKSCDKTHCPPCPAPPELLG 417
Db 180 -WFTNK--VYTCQVSHQG-VTQSRNI-----TSGQVPLSCND---P 213
QY 418 SVFLPPKPKDTLMISRTPEVTCVAVVDVSHEDVEKFNWYVDGVEVNAKTKPREEQYN 477
Db 214 VIKLPLPSIQVL-LEATVTLTCV---VSNAPGVNVMS---TQEQKSLSEIIVQGED 266
QY 478 TYRVSVTLVLHQMNLNGEKYKCVSNKALPAPLEKTSKAKGP-REPOV-YTLPPSRD 535
Db 267 ADSVISTVNISTQAMLSGAEFYCVNHDPLPTPLRASIHKEEVADLREPSVILLSPAED 326
QY 536 ELTNGVSLTCLVKGFPSPDIIVAVMESNGQPEN--NYKTPPVLDSD-GSPFLYSKLT 592
Db 327 VSAQRFSLTCLVKGFPSPDIIVAVMESNGQPEN--NYKTPPVLDSD-GSPFLYSKLT 592
QY 593 KSRWQGNVFCSVMEALNHNHYTKQKSLSPGLQDLECAEADQDGLDGLMTTD 647
Db 387 ABEWASGASVSCVGHAI-----PLKTIINRTYKSSDSS-DHIMIED 428
```

RESULT 90



S09265  
Ig alpha chain C region - rabbit (fragment)  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 16-Jul-1999  
C:Accession: S09265  
R:Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.  
EMBO J. 8, 4041-4047, 1989  
A:Title: The Iga heavy-chain gene family in rabbit: cloning and sequence analysis of 13  
A:Reference number: S09264; PMID:90076124; PMID:2512120  
A:Accession: S09265  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-357 <BUR>  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: immunoglobulin  
F:1246-319/Domain: immunoglobulin homology <IMM>

Query Match 8.6%; Score 293.5; DB 2; Length 357;  
Best Local Similarity 32.2%; Pred. No. 3.1e-10;  
Matches 89; Conservative 41; Mismatches 117; Indels 29; Gaps 12;

QY 369 OCLLSGQVLLSNIKVLPTWSTPVEPKSCDK-----THTCP-PCFAPELLGSPSVF 420  
DB 73 QCPBENSACRVEHNKRGDGV--TVPSPPACNESTIEPTPTPCPCPSPS-CGRPSLS 129  
QY 421 LPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNNYVGDVEVNAKTKPREE---QVNS 477  
DB 130 LQRPDLGD-LIANSNLSLTCLRGILNPPGAV-FTW----EPTGKEFVQGSPLDHC 182  
QY 478 TYRVSVLTVLHODWLNGEKYCKVSNKALPAP-PIEKTISKAKGQPREPOVYTLPPSRDE 536  
DB 183 CVSVSVLTPGCAVLNAGTEFTCTVTHPEIGDSLGTISKDTGSLIPQVHLPPSPSE 242  
QY 537 LKNG-VSLTCLVKGFYPSDIAVEMESNGO---PENNYKTTTPVLD---SDSFLYSKLT 589  
DB 243 LALNLVLTCLVRGSPFDVLSWTHNGTPPVVFKPSYLVWKRLEPPGDDPTTVAITSL 302  
QY 590 TVDKSRMOQGNVFCSVNHEALNHNHYTOKSLSPG 625  
DB 303 RVPEDMNGDSYSCVGHGELAHETQXTIDRLAG 338

RESULT 91  
HYRKS  
Ig mu chain C region, secreted (clone 3050) - horn shark  
C:Species: Heterodontus francisci (horn shark)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Jul-1999  
C:Accession: S01853  
R:Kokubu, F.; Hinder, K.; Litman, R.; Shambloct, M.J.; Litman, G.W.  
EMBO J. 7, 1979-1988, 1988  
A:Title: Complete structure and organization of immunoglobulin heavy chain constant region  
A:Reference number: S00980; PMID:85328985; PMID:3138109  
A:Accession: S01853  
A:Molecule type: DNA  
A:Residues: 1-438 <KOK>  
A:Cross-references: EMBL:X07781  
A:Note: the sequence was determined from the germ-line gene  
C:Genetics:  
A:Intron: 100/1; 206/1; 309/1  
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as Iga and Igm, the subunits associate into 16 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglobulin  
F:1-438/Domain: C region <CRE>  
F:120-87/Domain: immunoglobulin homology <IM1>  
F:123-190/Domain: immunoglobulin homology <IM2>  
F:228-291/Domain: immunoglobulin homology <IM3>  
F:330-400/Domain: immunoglobulin homology <IM4>  
F:164,200,245,275,374,411,415,425/Binding site: carbohydrate (Asn) (covalent) #status pr

Query Match 8.6%; Score 293.5; DB 1; Length 438;  
Best Local Similarity 25.8%; Pred. No. 4.1e-10;  
Matches 117; Conservative 68; Mismatches 163; Indels 105; Gaps 22;

QY 208 SSIYKKEGEQVEF---SFPLAFT-----VEKLTSGELMWOAKRASSSKSM 251  
DB 37 TSVSNKONERTITGLTKYPSVLANKGTYTQSSQTLTIESVSGSKITCEVRGSS--VW 94  
QY 252 ITFDLKNKEVSKRVTDPKLQMGKPLHLTLPALPOYAGSGLTTLALEA----- 303  
DB 95 I-----KEI-----PDCK---GDKHPVITLTQSSSEETSRFATVGLSIDFHES 139  
QY 304 -----KTKGKHQEVNLYVMATQLOKVLTEWGPSPKMLSLKENKAKSKREK 357  
DB 140 ITVSNLKGQ--HMSGFVT-----SPTGVNGTFSATSLT-----VARE-- 179  
QY 358 VVNLPEAGMMQCLSDSGQVLLSNIKVLPTWSTPVEPKSCDKHTCPAPPELLGSP 417  
DB 180 -MFTK---VITQVSHQG-VTQSNLT-----TGQVPPSCND----P 213  
QY 418 SVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKFNNYVGDVEVNAKTKPREEQVNS 477  
DB 214 VIKLPPSIEQL-LEATVTLTCLV---VSNAPYGVNWSM---TQOKSLKSEIAVQPED 266  
QY 478 TYRVSVLTVLHODWLNGEKYCKVSNKALPAP-PIEKTISKAKGQ-REPOV-YTLPPSRD 535  
DB 267 ADSVISTVNIQTQAWLSGAEPYCVVNHQDLPTPLRASIKKEVXDLREPSVILSPABD 326  
QY 536 ELTKNQVSLTCLVKGFYPSDIAVEMESNGOEN--NYKTTTPVLDSD--GSFLYSKLTVD 592  
DB 327 VSAQFLSLTCLVRKFFREIFVKWTVDKSYNGENKTEVMAENDNSYFTYLSLSIA 386  
QY 593 KSRMOQGNVFCSVNHEALNHNHYTOKSLSPG 625  
DB 387 AEEMASGASYSYCVGHEALPLKIIRTVNKSNG 419

RESULT 92  
S09274  
Ig alpha chain C region - rabbit (fragment)  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 16-Jul-1999  
C:Accession: S09274  
R:Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.  
EMBO J. 8, 4041-4047, 1989  
A:Title: The Iga heavy-chain gene family in rabbit: cloning and sequence analysis of 13  
A:Reference number: S09264; PMID:90076124; PMID:2512120  
A:Accession: S09274  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-347 <BUR>  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: immunoglobulin  
F:132-198/Domain: immunoglobulin homology <IMM>

Query Match 8.6%; Score 292.5; DB 2; Length 347;  
Best Local Similarity 32.3%; Pred. No. 3.5e-10;  
Matches 86; Conservative 40; Mismatches 117; Indels 23; Gaps 13;

QY 369 OCLLSGQVLLSNIKVLPTWSTPVEPKSCDKHTCPAPPELLGSPSVFLPPPKED 428  
DB 77 QCLEYDSAAACHVEYVSVINBSLPVF--PDPEQCH-CPSCRE-----PSLSLQRPDLRD 128  
QY 429 TLMISRTPEVTCVVVDVSHEDPE-VKFWV-VVDGVEVNAKTKPREEQVNSTYRVSVLT 486  
DB 129 -LLLSGDSASLTCLRGIL--KDPGAVLWGPFTNGNE--PVOQSPQRDC-GCYSVSVLP 182  
QY 487 VLHODWLNGEKYCKVSNKALP-PIEKTISKAKGQPREPOVYTLPPSRDELTKNQ-VSL 544  
DB 183 GCAEPMAAGTEFTCTVTHPEIGSSLTATISKDGLSLPRLVHLPPSEELALNALVT 242  
QY 545 TCLVKGFPSPDIAVEMESNG--QPBNNY---KTTTPVLDSDSGSFLYSKLTVDKSRMOQG 599  
DB 243 TCLVRGFPSPKVLVSWTKGKGVKVPENSEFLVWKRLEPPGDDPTTVAVSLRVPEDMNQ 302  
QY 600 NVFCSVNHGALNHNHYTOKSLSPG 625



Db 303 ESXSCVVAHEGLAEHFTQRTIDRLAG 328

## RESULT 93

Query Match 8.4%; Score 285.5; DB 2; Length 343;  
 Best Local Similarity 33.9%; Pred. No. 8.9e-10;  
 Matches 81; Conservative 35; Mismatches 100; Indels 23; Gaps 12;

WCHH  
 Ig mu chain C region - chicken (fragment)  
 C/Species: Gallus gallus (chicken)  
 C/Date: 03-Aug-1984 #sequence\_revision 03-Aug-1984 #text\_change 22-Jun-1999  
 C/Accession: A02170  
 R/Dahan, A.; Reynaud, C.A.; Weill, J.C.  
 Nucleic Acids Res. 11, 5381-5389, 1983  
 A/Title: Nucleotide sequence of the constant region of a chicken mu heavy chain immunoglobulin  
 A/Reference number: A02170; MUID:83299221; PMID:6310496  
 A/Accession: A02170  
 A/Molecule type: mRNA  
 A/Residues: 1-367 <DAH>  
 A/Cross-references: GB:K00389; NID:g212204; PIDN:AAA48923.1; PID:g212205  
 C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C/Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin  
 F/49-133/Domain: immunoglobulin homology <IMM2>  
 F/151-219/Domain: immunoglobulin homology <IMM3>  
 F/158-329/Domain: immunoglobulin homology <IMM4>  
 F/347-367/Domain: carboxyl-terminal <CTS>  
 F/51.119.303.354/binding site: carbohydrate (Asn) (covalent) #status predicted  
 F/156-111.158-217.265-327/Disulfide bonds: #status predicted  
 F/126.366/Disulfide bonds: interchain (to heavy chain) #status predicted  
 F/205/Disulfide bonds: interchain (to mu chain in another subunit) #status predicted

Query Match 8.5%; Score 291; DB 1; Length 367;  
 Best Local Similarity 30.8%; Pred. No. 4.6e-10;  
 Matches 91; Conservative 48; Mismatches 126; Indels 30; Gaps 14;

Qy 354 REKP--VWVUN--PEAGMOCCLSDSQVLESNIKLP--WSTPV-----EPKSC 399  
 Db 61 RRRPTEVWYKNGSGVAAATATATVGEVVAESRISTESEMPTGATFSCVGEKEMNT 120  
 Qy 400 DKHTHC--PCGPAPELLGSPVFLPPPKDITMISRPETCVVVDVSHED--EVKFNW 456  
 Db 121 SKRMECGLEPVQODI---AIRYTPSFVD-IFISKSATLTICRYSNNVNDGLEVSW-W 174  
 Qy 457 YVDGEVNAKTKPREEOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTIS 516  
 Db 175 KEKGKULETALGK-RVLOSNGLYTDGVAITCASBMDGDDYCKVNHPPDLLFMEEMKR 233  
 Qy 517 KAK-GQPREPOVYTLPSRDELTKNQ-VSLTCLVKGFYPSDIAVEWESNGQ--PENNYKT 572  
 Db 234 KTKASNAAPPSVYVFPPTTEQLNGNQRLSVTCMAQGFNPPLFVRMMKNGEPLPQSGSVT 293  
 Qy 573 TRPVLDSDGSFPLYSKLTVDKSRMOQGNVSCSVMEALHNHYTOKSLSLSPG 625  
 Db 294 SAPMAENENESYVAYSVLGVGAEEMGAGNVYTCLVGHEALPLQLAQKSVDRASG 348

## RESULT 94

S09272  
 Ig alpha chain C region - rabbit (fragment)  
 C/Species: Oryctolagus cuniculus (domestic rabbit)  
 C/Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 16-Jul-1999  
 C/Accession: S09272  
 R/Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.  
 EMBO J. 8, 4041-4047, 1989  
 A/Title: The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13 A/Reference number: S09264; MUID:90076124; PMID:2512120  
 A/Accession: S09272  
 A/Status: not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-343 <BUR>  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C/Keywords: immunoglobulin  
 F/232-305/Domain: immunoglobulin homology <IMM>

Query Match 8.4%; Score 285.5; DB 2; Length 343;  
 Best Local Similarity 33.9%; Pred. No. 8.9e-10;  
 Matches 81; Conservative 35; Mismatches 100; Indels 23; Gaps 12;

Qy 399 CDKHTT-CPPCAPBELIGSPVFLPPPKDITMISRPETCVVVDVSHEDPE-VKFNW 456  
 Db 97 CDPCHCYCPPTSC---GEPSSLORPDIIDLLESKA-SLTCTLSGL--KDPGAVFTW 149  
 Qy 457 Y-VDGEVNAKTKPREEOYN-STYRVSVLTVLHODMLNGEKYCKVSNKALP-APIEK 513  
 Db 150 EPTNG---NEFVQGSVQSTPPCGISVSVLPCCAEPMNAGTEFTCTVTHPEIGSLTA 205  
 Qy 514 TISKAKQPREPOVYTLPSRDELTKNQ-QVSLTCLVKGFYPSDIAVEWESNGQ--PENN 569  
 Db 206 TIGSRGSLTPPVYHLPPTEELALNEQVTLCTLVGFGSPKQVLYSWTHNGTLVVPKDS 265  
 Qy 570 Y---KTPPVLDSDGSFPLYSKLTVDKSRMOQGNVSCSVMEALHNHYTOKSLSLSPG 625  
 Db 266 YLWVKPLPEPGDPTVAVATSLRVSAMDNQSDSYSCVVGHGGLAEHFTQRTIDRQAG 324

## RESULT 95

S12328  
 Ig heavy chain C region (clone 5301) - horn shark (fragment)  
 C/Species: Heterodontus francisci (horn shark)  
 C/Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 21-Jan-2000  
 C/Accession: S12328; S01855  
 R/Litman, R.  
 Submitted to the EMBL Data Library, May 1988  
 A/Reference number: S12328  
 A/Accession: S12328  
 A/Molecule type: mRNA  
 A/Residues: 1-244 <LIT>  
 A/Cross-references: EMBL:X07785; NID:g633965; PIDN:CAA30618.1; PID:g633966  
 A/Note: this sequence was determined from the differentiated gene  
 R/Kokubu, F.; Hinds, K.; Litman, R.; Shambloct, M.C.; Litman, G.W.  
 EMBO J. 7, 1979-1988, 1988  
 A/Title: Complete structure and organization of immunoglobulin heavy chain constant region  
 A/Reference number: S00980; MUID:8328985; PMID:3138109  
 A/Accession: S01855  
 A/Molecule type: mRNA  
 A/Residues: 200-244 <KOK>  
 A/Cross-references: EMBL:X07785  
 A/Note: this sequence was determined from the differentiated gene  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C/Keywords: glycoprotein; heterotetramer; immunoglobulin; transmembrane protein  
 F/1-244/Domain: C region (fragment) <CRE>  
 F/111-181/Domain: immunoglobulin homology <IMM>  
 F/215-241/Domain: transmembrane #status predicted <TMN>  
 F/26.155.192.196/binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.3%; Score 284; DB 2; Length 244;  
 Best Local Similarity 32.2%; Pred. No. 7e-10;  
 Matches 75; Conservative 39; Mismatches 89; Indels 30; Gaps 9;

Qy 428 DTLMISRPETCVVVDVSHEDPEVKFNWYVDGEVNAKTKPREOY-----NSTYRVV 482  
 Db 4 EQVILEATVITLCV--VSNAPYGVANVSW-----TQOKPLKSIANVQPGSDSVI 52  
 Qy 483 SVTLVLHODMLNGEKYCKVSNKALPAPIEKTISKAGQP-REPOV-YTLPPSRDELTKN 540  
 Db 53 STYDISAQSWLSGDVFCVVSQDLPFLRDFIHKEKNKDLREBSVVLPPADVGAQR 112  
 Qy 541 QVSLTCLVKGFYPSDIAVEWESNGQPN--NYKTPPVLDSD--GSFPLYSKLTVDKSRMO 597  
 Db 113 FLSTLCVLRGFSRREIFVKTVNDKSVNPNGNVNTETVMAANDSSYFIYSLSLIADEWA 172  
 Qy 598 QGNVSCSVMEALH-----NHVYOKSLSLSPGLDDETCABAOGEGLDGLMTT 646  
 Db 173 SGASYSCLVGHGHALPKTIIRTVNKSDDSS-----DHVIEDNEESGNITWT 220



```

RESULT 96
B23360
Ig alpha-2 chain C region (allotype A2m(1)) - human
C/Species: Homo sapiens (man)
C/Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
C/Accession: B23360
R/Flanagan, J.G.; Lefranc, M.P.; Rabbitts, T.H.
C/11 36, 681-688, 1984
A/Title: Mechanisms of divergence and convergence of the human immunoglobulin alpha-1 ar
A/Reference number: A94653; MUID:84130179; PMID:6421489
A/Accession: B23360
A/Molecule type: DNA
A/Residues: 1-340 <FLA>
C/Genetics:
A/Gene: GDB:IGHA2
A/Cross-references: GDB:119333; OMIM:147000
A/Introns: 1/1 103/1; 210/1
C/Superfamily: Immunoglobulin C region; immunoglobulin homology
C/Keywords: immunoglobulin
F/230-302/Domain: immunoglobulin homology <IMM>

Query Match      8.3%; Score 282.5; DB 2; Length 340;
Best Local Similarity 28.3%; Pred. No. 1.3e-09;
Matches 100; Conservative 47; Mismatches 137; Indels 69; Gaps 17;

QY 333 PTSPKLM-LSLKLENKAKVSKR-----EKPVVVLNPEAGM-----WQCLLSDSQ 377
DB 3 PTSPKFLSLDSTQDGNVAVCLVQGFPOEPLSVTWSGQVTAANFPSPDASGD 62
QY 378 VLLSNIKVLPWTSPVEPKS--CD-KTHTCP-----PCPAPELLGSPVLFPPKP-- 426
DB 63 LYTTSSQLTLPATNQC-DEKSVTCHVKYHTNSQDVTVCPV-----PPPC 110
QY 427 -----KDTLMSRPEVTCVVVVDVSHEDPEVKFNWYDGVANAKTKPREBOYN 476
DB 111 HRLRLHRLPALDILLGSEANLCTLTGL-RDASGATFTWTPSSGK--SAVGPPPERDL 167
QY 477 STYRVSVLTVLHODMLNGKEYKCVSNKALPAPIEKITSKAKGQPREPOVTLPSRDE 536
DB 168 GCTSVSVLPSCAQPMNHETFTCTAAHPELKTPLTANITKS-GNTFREVLHLPSPSE 226
QY 537 LTKNO-VSLTCLVKGFPSDIAVENESNQ--PENNYKTPPVLD-SDG--SFPLYSKLT 590
DB 227 LALNELVLTCLARGPSKPDVLRHLQSQELPREKYLTMASROPSQTTFAATSLIR 286
QY 551 VDKSMQCGNVSCSVMEALHNHYTQKSLSPG-----LQIDETC 632
DB 287 VAABDMKGDTPSCWGHLEPLAFTQKTRIDLAGKPTHVNVSVMAEVDGTC 339

RESULT 97
HYRKC1
Ig mu chain C region (clone 6121) - horn shark (fragment)
C/Species: Heterodontus francisci (horn shark)
C/Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999
C/Accession: S01852
R/Kokubu, F.; Hinde, K.; Litman, R.; Shambloot, M.J.; Litman, G.W.
EMBO J. 7, 1979-1988, 1988
A/Title: Complete structure and organization of immunoglobulin heavy chain constant regi
A/Reference number: S00980; MUID:88328985; PMID:3138109
A/Accession: S01852
A/Molecule type: mRNA
A/Residues: 1-393 <KOK>
A/Cross-references: EMBL:X07782; NID:963961; PIDN:CAA30615.1; PID:9833623
A/Note: The sequence was determined from the differentially expressed gene
C/Complex: An immunoglobulin heterodimer subunit consists of two identical light (kap
chain) and heavy (kappa) chains. In some cases, such as IgA and IgM, the subunits associate into 16
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: duplication; glycoprotein; heterodimer; immunoglobulin
F/1-393/Domain: C region (fragment) <CRB>
F/1-42/Domain: immunoglobulin homology (fragment) <IM1>
F/78-145/Domain: immunoglobulin homology <IM2>

```

```

F/183-246/Domain: immunoglobulin homology <IM3>
F/285-355/Domain: immunoglobulin homology <IM4>
F/119,155,200,230,329,366,370,380/Binding site: carbohydrate (Asn) (covalent) #status pr

Query Match      8.2%; Score 281.5; DB 1; Length 393;
Best Local Similarity 26.7%; Pred. No. 1.8e-09;
Matches 116; Conservative 63; Mismatches 164; Indels 91; Gaps 22;

QY 211 VYKKEGEVSEFPFLAFVTEKLTGSGELMQWQERASSKSWTFPLKNKEVSKRVTD 270
DB 13 VLNKGTYTQ--SQGLTIT-ESEVSSKTYCEVRGES--VLI-----PDC 56
QY 271 KLQMGKULPLHLTLPQALPOYAGSGNLTALAE-----XTGKLHGVNLV 316
DB 57 K--GDVHPVYLTLQSSSEITSRFAVLCIIDHFEISTVSWLMDGQ-HMESGVT 112
QY 317 MRATQQLNLTCEWGPSPKLMLSLKENKAKVSKREKPVVVLNPEAGMQLLSDSG 376
DB 113 -----SPICGVNGETSATSRLT-----VPAE--MFTNK--VYTCQVSHQ 149
QY 377 QVLESNIKVLPTWTPVEPKSCDKTHTCPCPAPELLGSPVLFPPKPDTLMSRTP 436
DB 150 -VTGSRNL-----TGSQVPCND-----PVIKLPLPSIRQVL-LEATV 186
QY 437 EYTCVVVDVSHEDPEVKENWYDGVANAKTKPREBOYNSTYRVSVLTVLHODMLNGK 496
DB 187 TLNCL--VSNAPYVNVSW---TQEQSLKSEIAVQGEDADSVISTVNIStOMLSGA 240
QY 497 EYCKVSNKALPAPIEKITSKAKGP-REPOV-YTLPSRDELTKNQVSLTCLVKGPYS 554
DB 241 EYFCVNVNODLPTPLRASIHKEVEVDLREPVSILSLPAEDVSAQRFLSLTCLVNGFAPR 300
QY 555 DIAVEMENQGPEN--NKITPPVLDSD-GSFPLYSKLTVDKSRQCGNVSCSVMEAL 611
DB 301 EIPVMTINDSVMPGNKNTVMAENDSSVFLYSLISIAAEWASGASVSCVGHBAI 360
QY 612 HNHVYQKSLSPG 625
DB 361 PLKIINRTVNSSG 374

RESULT 98
S21462
T-cell surface glycoprotein CD4 (allele 2) - pig (fragment)
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 21-Jan-2000
C/Accession: 147132; S21462
R/Gustafsson, K.; Germana, S.; Sundt, T.M.
J. Immunol. 151, 1365-1370, 1993
A/Title: Extensive allelic polymorphism in the CDR2-like region of the miniature swine C
A/Reference number: 147131; MUID:93329116; PMID:8335933
A/Accession: 147132
A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-99 <GU2>
A/Cross-references: EMBL:X65630; NID:91929; PIDN:CAA46584.1; PID:9388233
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C/Keywords: glycoprotein; T-cell
F/3-81/Domain: immunoglobulin homology <IMM>

Query Match      8.2%; Score 280.5; DB 2; Length 99;
Best Local Similarity 56.1%; Pred. No. 3.6e-10;
Matches 55; Conservative 19; Mismatches 23; Indels 1; Gaps 1;

QY 32 KKGDTVELTCTASQKKSIOFMKNSNQIKILNGGSPFLTKGP-SKLNDRADRSRLMDG 90
DB 1 KAGDLAELPCHSSQKQKLPFWKNSDQKILSRHNLNWKASVTELSRLDSKKQMDHG 60
QY 91 NFPLIKLKLESDTYICEVEDQKEVQLLVGLTAN 128
DB 61 SFPLIKKLEVTDSGIYICEVEDKRIEVOGLVFLTLAS 98

```



```

RESULT 99
Ig alpha chain C region - rabbit (fragment)
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 16-Jul-1999
C/Accession: S09266
R/Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.
EMBO J. 8, 4041-4047, 1989
A/Title: The IGA heavy-chain gene family in rabbit: cloning and sequence analysis of 13
A/Reference number: S09264; MUID:90076124; PMID:2512120
A/Accession: S09266
A/Status: not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-352 <BUR>
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: immunoglobulin
F/241-314/Domain: immunoglobulin homology <IMM>

Query Match      8.2%; Score 280.5; DB 2; Length 352;
Best Local Similarity 32.2%; Pred. No. 1.8e-09;
Matches 85; Conservative 37; Mismatches 109; Indels 33; Gaps 13;

Qy 374 DSGVLLSNIKVLPTWSTPEKSCDKHTCPCPAPDEL-LGQPSVFLPFPKPKDTLMI 432
Db 91 DEGNL-----TVLYPECKDPNSDPTCPCPPTCGEPSTLQRPDID-LLL 137
Qy 433 SRTEVTCVVVDVSHEDPE-VKFNMY-VDGVEVFNNAKTKREBEQYN-STYVVSVLTVMH 489
Db 138 ESNASLTTLTSLG-L-KDEGAIVFTWNPNTNGNEPVQOQST---QSPCCGVSVSVLPQCA 191
Qy 490 QDWLNKKEYKCKVSNKALP-APIEKTSKAKGQPREPQVYTLPPSRDELTKNQ-VSLTCL 547
Db 192 EPMNAGIEFTCTVHPELGGSLATISRSGLTP--PQHLLPPTPEELANALVTILCL 249
Qy 548 VKGPPSPDIAVESNSGQ--PENNY--KTPPYLSDSGSFYLSKLTVDKSRMQGNV 601
Db 250 VRGSPKVLVSWTHNGTLVLPKPSFLWKPLPEPGCEPTTAAVTSILRLVAEDMNOGDS 309
Qy 602 FSGVMHEALHNHYTQKSLSLSPG 625
Db 310 YSCVGHGEGALAEHFTQRTIDRLAG 333

RESULT 100
A/HU
Ig alpha-1 chain C region - human
C/Species: Homo sapiens (man)
C/Date: 22-May-1991 #sequence_revision 03-Oct-1995 #text_change 20-Oct-2000
C/Accession: A22360; A92249; A91662; S38979; B53110; A02171
R/Flanagan, J.G.; Lefranc, M.P.; Rabbitts, T.H.
Cell 36, 681-688, 1984
A/Title: Mechanisms of divergence and convergence of the human immunoglobulin alpha-1 at
A/Reference number: A94653; MUID:84130179; PMID:6421489
A/Accession: A22360
A/Molecule type: DNA
A/Residues: 1-353 <FLA>
R/Putnam, F.W.; Liu, Y.S.V.; Low, T.L.K.
J. Biol. Chem. 254, 2865-2874, 1979
A/Title: Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal IgA protease
A/Reference number: A92249; MUID:79151016; PMID:107164
A/Contents: myeloma protein Bur; disulfide bonds
A/Accession: A92249
A/Molecule type: protein
A/Residues: 1-16, 'Z', 18, 'B', 20, 'B', 22-34, 'Q', 36-45, 'Z', 47-51, 'B', 53-56, 'ZB', 59-61, 'B', 63
303, 'B', 305-346, 'Q', 348-353 <PUT>
A/Note: this is the final paper in a series
R/Kratz, H.; Alevoigt, P.; Ruban, E.; Kortt, A.; Stareschik, K.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 356, 1337-1342, 1975
A/Title: The primary structure of a monoclonal IgA-immunoglobulin (IgA Tro.), II: the an
A/Reference number: A91662; MUID:76023781; PMID:809331
A/Accession: A91662
A/Molecule type: protein

```

```

A/Residues: 1-16, 'Z', 18, 'B', 20, 'B', 22-34, 'Q', 36-45, 'Z', 47-51, 'B', 53-56, 'ZB', 59-61, 'B', 63-
'R', 232-237, 'Q', 240-243, 'Q', 245-283, 'Q', 285-289, 'E', 291-303, 'B', 305-353 <KRA>
A/Experimental source: myeloma protein Tro
R/Falgreen-Gebauer, E.; Gebauer, W.; Baettin, A.; Kratz, H.D.; Eifert, H.; Zimmermann
Biol. Chem. Hoppe-Seyler 374, 1023-1028, 1993
A/Title: The covalent linkage of secretory component to IgA. Structure of sIgA.
A/Reference number: S38978; MUID:94121784; PMID:8292260
A/Accession: S38978
A/Molecule type: protein
A/Residues: 188-196, 'D', 198-201 <PAL>
R/Tang, C.Y.; Kratz, H.; Gocz, H.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 360, 1919-1940, 1979
A/Title: Die Primärstruktur eines monoklonalen IgA1-Immunglobulins (Myelomprotein Tro) .
A/Reference number: A91684; MUID:80114124; PMID:393607
A/Contents: annotation; Tro; disulfide bonds
A/Note: Cys-14 bonds to a light chain
R/Calejo, M.; Bescibes, U.; Grubb, A.; Mendez, E.
J. Biol. Chem. 269, 384-389, 1994
A/Title: Location of a novel type of interpeptide chain linkage in the human protein
A/Reference number: A53110; MUID:94103241; PMID:7506257
A/Accession: B53110
A/Molecule type: protein
A/Residues: 346-351, 'X', 353 <CAL>
C/Genetic:
A/Gene: GDB:IGHA1
A/Map position: 14q32.33-14q32.33
A/Intons: 1/1; 103/1; 223/1
C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into lar
C/Superfamily: immunoglobulin C region; immunoglobulin homology
C/Keywords: chromoprotein; duplication; glycoprotein; heterotetramer; immunoglobulin; tr
F/140-206/Domain: immunoglobulin homology <IM1>
F/243-315/Domain: immunoglobulin homology <IM2>
F/26-85, 77-101, 123-180, 147-204, 250-313/Disulfide bonds: #status experimental
F/105, 111, 113, 119, 121/Binding site: carbohydrate (Ser) (covalent) #status experimental
F/122, 182/Disulfide bonds: interchain #status experimental
F/144, 340/Binding site: carbohydrate (Asn) (covalent) #status experimental
F/192/Binding site: interchain (to secretory component) (partial) #status experimental
F/192/Binding site: cysteine (Cys) (covalent) (partial) #status experimental
F/192/Disulfide bonds: interchain (partial) #status experimental
F/352/Cross-link: alpha-1-microglobulin-Ig alpha complex chromophore (Cys) (interchain to
F/352/Disulfide bonds: interchain (to J chain) (partial) #status experimental
F/352/Disulfide bonds: interchain (partial) #status experimental

Query Match      8.2%; Score 280.5; DB 1; Length 353;
Best Local Similarity 26.8%; Pred. No. 1.8e-09;
Matches 93; Conservative 49; Mismatches 132; Indels 73; Gaps 13;

Qy 305 TGKLGQVNLVWRATQ--LQKNLTCEVWGPTSPKMLSLKLNKAKVSKREKPVVVLN 362
Db 60 SGDLTYTSSQULPAPDCLAGKSVTCVVKHYTPSQDVTY----- 100
Qy 363 PEAGMOCILSDSGVLLSNIKVLPTWSTPEKSCDKHTCPCPAPDELGGPSVFLF 422
Db 101 -----CPVSTRPPTP-----SPSTRPPTSPSCC-HPRLSLH 130
Qy 423 PPKPKDTLMSRPETVCVVVDVSHEDPEVKFNMYVDGVVNAKTPREQNSTRVV 482
Db 131 RPALED-LILGSEBANILCTLTGL-RDASGTFWTBSSGR--SAVQPPRRDLGCGSVS 186
Qy 483 SVLTVHODLNGKEYKCKVSNKALPAPIEKTSKAKGQPREPQVYTLPPSRDELTKNQ- 541
Db 187 SVLPQCEPNNHGKTFCTAAYPESKTPLATLTSKS-GNTFRREVLLPPESEELANEL 245
Qy 542 VSLTCLVKGPPSPDIAVESNSGQ--PENNYKTPPYLSD--SG--SFFLYSKLTVDKSRW 596
Db 246 VTILCLARGSPRDVLRWLGSGQELPREKYLTMASQDESGQTTTAAVLSILRLVAEDW 305
Qy 597 QQGNVPSGVMEALHNHYTQKSLSLSPG-----LQIDETC 632
Db 306 KKGDTPSCWVGHGALPLAFQKTIIDRLAGKPTVNVSVVAEVDGTC 352

```



Search completed: August 3, 2004, 13:14:49  
Job time : 20.8924 secs

---



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:44 ; Search time 8.8991 Seconds  
(without alignments)  
3791.557 Million cell updates/sec

Title: SEQ3  
Perfect score: 3414  
Sequence: 1 MNRGVPRHLLVLQLALLP.....DETCAADGDELDTTDP 648

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 125 summaries

Database : SwIsEProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2024	59.3	458	1	CD4_HUMAN
2	2000	58.6	458	1	CD4_PANTR
3	1853	54.3	458	1	CD4_MACFU
4	1852	54.2	458	1	CD4_MACMU
5	1849	54.2	458	1	CD4_MACFA
6	1844	54.0	458	1	CD4_MACNE
7	1827	53.5	458	1	CD4_CERAE
8	1726	50.6	397	1	CD4_ERYPA
9	1716	50.3	397	1	CD4_CERTO
10	1581.5	46.3	457	1	CD4_SATSC
11	1266	37.1	330	1	GCI_HUMAN
12	1151	33.7	290	1	GCI_HUMAN
13	1150	33.7	326	1	GCI_HUMAN
14	1149	33.7	459	1	CD4_RABIT
15	1141	33.4	327	1	GCI_HUMAN
16	1141	33.4	463	1	CD4_CANFA
17	999	29.3	457	1	CD4_RAT
18	993	29.1	457	1	CD4_MOUSE
19	947.5	27.8	398	1	GCI_MOUSE
20	940.5	27.5	393	1	GCI_MOUSE
21	922	27.0	323	1	GCI_MOUSE
22	915.5	26.8	399	1	GCI_MOUSE
23	906.5	26.6	329	1	GCI_MOUSE
24	884.5	25.9	405	1	GCI_MOUSE
25	849.5	24.9	329	1	GCI_MOUSE
26	844	24.7	333	1	GCI_MOUSE
27	828	24.3	326	1	GCI_MOUSE
28	826.5	24.2	326	1	GCI_MOUSE
29	817.5	23.9	329	1	GCI_MOUSE
30	816	23.9	335	1	GCI_MOUSE
31	814.5	23.9	330	1	GCI_MOUSE
32	795	23.3	322	1	GCI_MOUSE
33	788.5	23.1	336	1	GCI_MOUSE
34	375.5	11.0	476	1	MUCM_MOUSE
35	364.5	10.7	421	1	EPC_MOUSE
36	364	10.7	455	1	MUC_MOUSE
37	359.5	10.5	454	1	MUC_HUMAN
38	359	10.5	428	1	EPC_HUMAN
39	358.5	10.5	479	1	MUCM_RABIT
40	355.5	10.4	429	1	EPC_RAT
41	355.5	10.4	454	1	MUC_MESAU
42	353.5	10.2	391	1	MUC_HUMAN
43	349	10.2	458	1	MUC_FABIT
44	348	10.2	457	1	MUC_SUNMU
45	340	10.0	450	1	MUC_CANFA
46	305	8.9	299	1	ALC_RABIT
47	298	8.7	446	1	MUC_CHICK
48	297	8.7	438	1	HVC2_HETTR
49	295	8.6	461	1	HVCN_HETTR
50	293.5	8.6	438	1	HVCN_HETTR
51	282.5	8.3	340	1	ALC2_HUMAN
52	281.5	8.2	393	1	HVC3_HETTR
53	280.5	8.2	353	1	ALC1_HUMAN
54	278	8.1	353	1	ALC1_GORGO
55	267	7.8	370	1	HVC1_HETTR
56	265.5	7.8	481	1	MUCM_HETTR
57	248.5	7.3	344	1	ALC_MOUSE
58	203.5	6.0	513	1	SHS1_MOUSE
59	187.5	5.5	506	1	SHS1_BOVIN
60	182.5	5.3	739	1	VCAI_RAT
61	178	5.2	4391	1	PGBM_HUMAN
62	177.5	5.2	105	1	UN89_MOUSE
63	174.5	5.1	6632	1	UN89_CAEBL
64	172	5.0	104	1	LAC2_RAT
65	168.5	4.9	3707	1	PGBM_MOUSE
66	165.5	4.8	1493	1	NROI_MOUSE
67	163	4.8	104	1	LAC3_MOUSE
68	162.5	4.8	213	1	IL11_HUMAN
69	162.5	4.8	847	1	CD22_HUMAN
70	161.5	4.7	739	1	VCAI_HUMAN
71	161	4.7	104	1	LAC2_MOUSE
72	161	4.7	258	1	HB2D_PIG
73	161	4.7	702	1	CEA5_HUMAN
74	159	4.7	105	1	LAC_HUMAN
75	156	4.6	105	1	LAC_PIG
76	155	4.5	1906	1	KML5_CHICK
77	154	4.5	103	1	LAC3_MOUSE
78	153.5	4.5	103	1	LAC3_MOUSE
79	153	4.5	105	1	LAC_RABIT
80	153	4.5	106	1	VCAI_MOUSE
81	152	4.5	739	1	VCAI_MOUSE
82	150	4.4	106	1	KACA_RAT
83	150	4.4	509	1	SHS1_RAT
84	149	4.4	261	1	HB2C_PIG
85	148.5	4.3	2012	1	DBCA_HUMAN
86	148	4.3	106	1	KAC_HUMAN
87	148	4.3	1447	1	DCC_MOUSE
88	147.5	4.3	503	1	SHS1_HUMAN
89	147	4.3	398	1	SHS1_HUMAN
90	147	4.3	6885	1	SNE2_HUMAN
91	146	4.3	104	1	LAC1_RAT
92	146	4.3	387	1	SRB2_HUMAN
93	145.5	4.3	103	1	LAC_CHICK
94	145	4.2	106	1	KACB_RAT
95	144.5	4.2	555	1	CL66_CARAU
96	144.5	4.2	1257	1	CAML_HUMAN
97	144.5	4.2	1260	1	CAML_MOUSE
98	143	4.2	106	1	KAC_MOUSE
99	142	4.2	837	1	NCM2_MOUSE
100	141.5	4.1	761	1	NCM2_HUMAN
101	141	4.1	1259	1	CAML_RAT
102	140	4.1	268	1	HB2X_HUMAN
103	140	4.1	273	1	DOB_HUMAN
104	139.5	4.1	1234	1	NPBN_RAT
105	139	4.1	273	1	DOB_PANTR
106	138.5	4.1	1197	1	CAML_BRAAE
107					P01873 mus musculus
108					P06336 mus musculus
109					P01872 mus musculus
110					P01871 mus musculus
111					P01854 mus musculus
112					P04221 oryctolagus
113					P01855 mus musculus
114					P06337 mesocricetus
115					P04220 mus musculus
116					P03988 oryctolagus
117					P07068 sunco murt
118					P01874 canis faml
119					P01879 oryctolagus
120					P01875 gallus gall
121					P23085 heterodontu
122					P23088 heterodontu
123					P23087 heterodontu
124					P01877 mus musculus
125					P23086 heterodontu
126					P01876 mus musculus
127					P20758 gorilla gor
128					P23084 heterodontu
129					P23735 ictalurus p
130					P01878 mus musculus
131					P97797 m protein-t
132					O46631 bos taurus
133					P28534 rattus norv
134					P98160 mus musculus
135					P01843 mus musculus
136					O01761 caenorhabd
137					P20767 rattus norv
138					O05793 mus musculus
139					P97798 mus musculus
140					P01845 mus musculus
141					P15814 mus musculus
142					P20273 mus musculus
143					P19320 mus musculus
144					P01844 mus musculus
145					P15983 mus musculus
146					P06731 mus musculus
147					P01842 mus musculus
148					P01846 mus musculus
149					P11799 gallus gall
150					P01880 mus musculus
151					P20765 mus musculus
152					P01847 oryctolagus
153					P01839 oryctolagus
154					P29533 mus musculus
155					P29536 mus musculus
156					P97710 r protein-t
157					P15982 mus musculus
158					O6469 mus musculus
159					P01834 mus musculus
160					P07211 mus musculus
161					P78324 h protein-t
162					O00241 mus musculus
163					O8wxn0 mus musculus
164					P20766 mus musculus
165					P01837 mus musculus
166					O35136 mus musculus
167					P13592 mus musculus
168					O06695 mus musculus
169					P05538 mus musculus
170					P13765 mus musculus
171					O9044 mus musculus
172					P18467 mus musculus
173					O90478 mus musculus



107	137.5	4.0	564	1	C166_BRARE	Q09460 brachydanio
108	136	4.0	261	1	HB24_HUMAN	P01920 homo sapien
109	136	4.0	997	1	SCPI_RAT	O03410 rattus norv
110	135.5	4.0	105	1	LAC5_MOUSE	P20764 mus musculu
111	135.5	4.0	848	1	NCAL_HUMAN	P13591 homo sapien
112	135.5	4.0	1447	1	DCG_HUMAN	P43146 homo sapien
113	134.5	3.9	261	1	HB22_HUMAN	P01919 homo sapien
114	134.5	3.9	1348	1	VGR2_COTVA	P52583 coturnix co
115	134	3.9	1369	1	NFAS_CHICK	O42414 gallus galli
116	133.5	3.9	1242	1	NPHN_MOUSE	O9qz57 mus musculu
117	132.5	3.9	261	1	HB21_HUMAN	P01918 homo sapien
118	132.5	3.9	723	1	NCAL_MOUSE	P13594 mus musculu
119	132	3.9	231	1	HB2L_CHICK	P23068 gallus galli
120	131.5	3.9	1240	1	NFAS_MOUSE	O810u3 mus musculu
121	131	3.8	261	1	HB23_HUMAN	P05537 homo sapien
122	130	3.8	837	1	NCM2_HUMAN	O15394 homo sapien
123	130	3.8	1302	1	NGG_DROME	P20241 drosophila
124	129.5	3.8	853	1	NCAL_BOVIN	P31836 bos taurus
125	128	3.7	1097	1	PCDR_RAT	Q05030 rattus norv

## ALIGNMENTS

RESULT 1  
CD4\_HUMAN STANDARD; PRT; 458 AA.

AC P01730;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen CD4, Leu-3).  
GN CD4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85254948; PubMed=2990730;  
RA Maddon P.J., Littman D.R., Godfrey M., Maddon D.E., Chese L., Axel R.;  
RT "The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin gene family.";  
RL Cell 42:93-104(1985).  
RN [2]  
RP REVISION TO 26.  
RX MEDLINE=89028665; PubMed=3263213;  
RA Littman D.R., Maddon P.J., Axel R.;  
RT "Corrected CD4 sequence.";  
RL Cell 55:541-541(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96303695; PubMed=8723724;  
RA Anaezi-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S., Malley T., Gibbs R.A.;  
RT "A gene-rich cluster between the CD4 and triosephosphate isomerase genes at human chromosome 12p13.";  
RL Genome Res. 6:314-326(1996).  
RN [4]  
RP SEQUENCE FROM N.A., AND VARIANT TRP-265.  
RX MEDLINE=91216786; PubMed=1708753;  
RA Hodge T.W., Sasso D.R., McDougal J.S.;  
RT "Humans with OKT4-epitope deficiency have a single nucleotide base change in the CD4 gene, resulting in substitution of TRP240 for ARG240.";  
RL Hum. Immunol. 30:99-104(1991).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas; PubMed=12477932;  
RX MEDLINE=22386257;  
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D., Altschul S.F., Zeeberg B., Bueto R.H., Schaefer C.F., Bhat N.K., RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P., RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L., RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schreier T.E., RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C., RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., RA Richards S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., RA Butlerfield Y.S.N., Krzywinski M.I., Skalka U., Smallus D.E., RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [6]  
RP SEQUENCE OF 28-424 FROM N.A.  
RC TISSUE=Blood;  
RX MEDLINE=93049640; PubMed=1425921;  
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;  
RT "Cloning and sequences of primate CD4 molecules: diversity of the cellular receptor for simian immunodeficiency virus/human immunodeficiency virus";  
RL Eur. J. Immunol. 22:2973-2981(1992).  
RN [7]  
RP SEQUENCE OF 26-394  
RX MEDLINE=90078233; PubMed=2592374;  
RA Carr S.A., Hemling M.E., Folena-Wasserman G., Sweet R.W., Anumula K., RA Barr J.R., Huddleston M.J., Taylor P.;  
RT "Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor by mass spectrometry";  
RL J. Biol. Chem. 264:21286-21295(1989).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 26-208.  
RX MEDLINE=91061881; PubMed=1701030;  
RA Wang J., Yan Y., Garrett T.P., Liu J., Rodgers D.W., Garlick R.L., RA Tarr G.E., Husain Y., Reinherz E.L., Harrison S.C.;  
RT "Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains";  
RL Nature 348:411-418(1990).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 26-208.  
RX MEDLINE=91061882; PubMed=2247146;  
RA Ryu S.-E., Kwong P.D., Truneh A., Porter T.G., Athos J., RA Rosenberg M., Dai X., Xuong N.-H., Axel R., Sweet R.W., RA Hendrickson W.A.;  
RT "Crystal structure of an HIV-binding recombinant fragment of human CD4";  
RL Nature 348:419-426(1990).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (3.9 ANGSTROMS) OF 26-388.  
RX MEDLINE=97311402; PubMed=9168119;  
RA Wu H., Kwong P.D., Hendrickson W.A.;  
RT "Dimeric association and segmental variability in the structure of human CD4";  
RL Nature 387:527-530(1997).  
RN [11]  
RP PALMITOYLATION.  
RX MEDLINE=92317088; PubMed=1618861;  
RA Crise B., Rose J.K.;  
RT "Identification of palmitoylation sites on CD4, the human immunodeficiency virus receptor";  
RL J. Biol. Chem. 267:13593-13597(1992).  
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regulate T-cell activation.  
CC -1- SUBUNIT: Associates with p56-lck.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.



CC	-I- DATABASE: NAME=PROW; NOTE=CD guide CD4 entry;
CC	WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd4.htm"
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M12807; AAA35572.1; -
DR	EMBL; U47924; AAB51509.1; -
DR	EMBL; M35160; AAAL6069.1; -
DR	EMBL; BC025782; AAH35782.1; -
DR	PIR; A90872; RMHTU4.
DR	PDB; 1CDH; 30-APR-94.
DR	PDB; 1CDI; 30-APR-94.
DR	PDB; 1CDJ; 31-OCT-93.
DR	PDB; 1CDU; 01-APR-97.
DR	PDB; 1CDY; 01-APR-97.
DR	PDB; 1CDB; 12-MAR-97.
DR	PDB; 1WIO; 07-JUL-97.
DR	PDB; 1WIQ; 07-JUL-97.
DR	PDB; 1G9M; 27-DEC-00.
DR	PDB; 1G9N; 27-DEC-00.
DR	PDB; 1GC1; 19-AUG-98.
DR	PDB; 1JL4; 19-SEP-01.
DR	GlycoSuiteDB; P01730; -
DR	GeneW; HGNC:1678; CD4.
DR	MIM; 186940; -
DR	GO; GO:0042101; C:T-cell receptor complex; NAS.
DR	GO; GO:0015026; F:coreceptor activity; NAS.
DR	GO; GO:0015029; F:internalization receptor activity; TAS.
DR	GO; GO:0042289; F:MHC class II protein binding; NAS.
DR	GO; GO:0004888; F:transmembrane receptor activity; TAS.
DR	GO; GO:0006955; P:immune response; NAS.
DR	GO; GO:0009405; P:pathogenesis; TAS.
DR	GO; GO:0030217; P:T-cell differentiation; NAS.
DR	GO; GO:0045058; P:T-cell selection; NAS.
DR	GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . ; NAS.
DR	InterPro; IPR000973; CD4 TCAG.
DR	InterPro; IPR002710; Ig-Like.
DR	InterPro; IPR003596; Ig_v.
DR	Pfam; PF00047; Ig_2.
DR	PRINTS; PR00692; CD4TCANTIGEN.
DR	SMART; SMO0406; IGV; 1.
DR	PROSITE; PS50835; IG_LIKE; 1.
KW	Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW	Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure;
KW	Polymorphism.
FT	SIGNAL 1 25
FT	CHAIN 26 458
FT	DOMAIN 26 396
FT	TRANSMEM 397 418
FT	DOMAIN 419 458
FT	DOMAIN 26 125
FT	DOMAIN 126 203
FT	DOMAIN 204 317
FT	DOMAIN 318 374
FT	CARBOHYD 296
FT	
FT	CARBOHYD 325 325
FT	
FT	DISULFID 41 109
FT	DISULFID 155 184
FT	LIPID 328 370
FT	LIPID 419 419
FT	LIPID 422 422
FT	VARIANT 265
FT	
CC	BY SIMILARITY.
CC	S-palmitoyl cysteine.
CC	R-palmitoyl cysteine.
CC	R -> W (in OKT4-negative populations).

FT	STRAND	27	32	
FT	TURN	33	34	
FT	STRAND	37	39	
FT	TURN	44	45	
FT	STRAND	51	55	
FT	TURN	56	57	
Query Match				
Best Local Similarity 59.3%, Score 2024; DB 1; Length 458;				
Matches 393; Conservative 1; Mismatches 2; Indels 0; Gaps 0;				
QY	1	MNRGVPFRHLHLVLTQALLPAAATGQKRVYLGAKKDTVELTCTASQKKSIGPFHMKNSQIK	60	
DB	1	MNRGVPRFHLHLVLTQALLPAAATGQKRVYLGAKKDTVELTCTASQKKSIGPFHMKNSQIK	60	
QY	61	ILGNQGSFLTTKPSPKLNDRADRSRLMDQGNFLLIKNLKIEDSDYTCEDQKEVQL	120	
DB	61	ILGNQGSFLTTKPSPKLNDRADRSRLMDQGNFLLIKNLKIEDSDYTCEDQKEEVQL	120	
QY	121	LVFGLTNSDTHLLQGGSLTTLTSPGSSPSVQCRPRGNIGQKTLVSQELDQSG	180	
DB	121	LVFGLTNSDTHLLQGGSLTTLTSPSSPSVQCRPRGNIGQKTLVSQELDQSG	180	
QY	181	TWTCTVLOQNKKEVFKDIIVIAFOKASSIYKKKEGQVEFSPLAFTVEKLTGSGELMW	240	
DB	181	TWTCTVLOQNKKEVFKDIIVIAFOKASSIYKKKEGQVEFSPLAFTVEKLTGSGELMW	240	
QY	241	QAEBASSKSWITFDLKNKEVSVKRVYQDPKLQWKKLPLHLTLPOALPOYAGSGNLTLLA	300	
DB	241	QAEBASSKSWITFDLKNKEVSVKRVYQDPKLQWKKLPLHLTLPOALPOYAGSGNLTLLA	300	
QY	301	LEATGKTLHQEVNIVVWRATQLOKLTCEVWGPTSPKLTMLSKLENKAKYSKREKPVVY	360	
DB	301	LEATGKTLHQEVNIVVWRATQLOKLTCEVWGPTSPKLTMLSKLENKAKYSKREKAVVY	360	
QY	361	LNPEAGMWQCLLSDSGOVLLESNIKVLPTWSTPVP	396	
DB	361	LNPEAGMWQCLLSDSGOVLLESNIKVLPTWSTPVP	396	
RESULT 2				
CD4_PANTR	STANDARD	PRT	458 AA	
AC	P16004			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-APR-1990 (Rel. 14, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	T-cell surface glycoprotein CD4 precursor (T-cell surface antigen CD4).			
GN	CD4.			
OS	Pan troglodytes (Chimpanzee).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.			
OX	NCBI_TaxID=9598;			
RN	[1]			
RX	SEQUENCE FROM N.A.			
RT	MEDLINE=90182664; PubMed=2107024;			
RA	Camerini D., Seed B.;			
RT	"A CD4 domain important for HIV-mediated syncytium formation lies outside the virus binding site.";			
RL	Cell 60:747-754(1990).			
RN	[2]			
RP	SEQUENCE OF 26-424 FROM N.A.			
RC	TISUB=Blood;			
RX	MEDLINE=93049640; PubMed=1425221;			
RA	Fomsgaard A., Hirsch V.M., Johnson P.R.;			
RT	"Cloning and sequences of primate CD4 molecules: diversity of the cellular receptor for simian immunodeficiency virus/human immunodeficiency virus.";			
RL	Eur. J. Immunol. 22:2973-2981(1992).			
CC	-1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regulate T-cell activation.			



```

CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31135; AAA5407.1; -.
DR EMBL; X73323; CAA51749.1; -.
DR PIR; B32722; RMCZT4.
DR HSSP; P01730; 1MR0.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR007110; Ig-Like.
DR InterPro: IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 1 25
FT DOMAIN 26 458
FT TRANSMM 397 458
FT DOMAIN 419 458
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 396
FT CARBOHYD 296 374
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT CONFLICT 42 62
FT CONFLICT 62 62
FT CONFLICT 191 191
SQ SEQUENCE 458 AA; 51057 MW; A7C3AC8A525D7D3AD CRC64;
Query Match 58.6%; Score 2000; DB 1; Length 458;
Best Local Similarity 98.2%; Pred. No. 4-7e-12;
Matches 389; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```

```

Db 181 TWCTCTVLONOKKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTVEKLTSSGELMW 240
Oy QAERASSKSMITFDLKNKKEVSRYVTPQDPLQWKKPLPLTLTPPLQPVYAGSGNTLTA 300
Db 241 QAERASSKSMITFDLKNKKEVSRYVTPQDPLQWKKPLPLTLTPPLQPVYAGSGNTLTA 300
Oy 301 LEAKTGKHOEVLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKARVSKREKPVW 360
Db 301 LEAKTGKHOEVLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKARVSKREKPVW 360
Oy 361 LNPAGMWOCLLSDSGVLLSESNIKVLPWTSTPVP 396
Db 361 LNPAGMWOCLLSDSGVLLSESNIKVLPWTSTPVP 396

RESULT 3
CD4_MACFU STANDARD; PRT; 458 AA.
ID CD4_MACFU
AC P79184;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OC NCBI_TaxID=9543;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatsumi M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63348; BAA09672.1; -.
DR HSSP; P01730; 1MR0.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR007110; Ig-Like.
DR InterPro: IPR003596; Ig_v.
DR Pfam; PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMM 397 418
FT DOMAIN 419 458

```



FT	DOMAIN	26	125	IG-LIKE V-TYPE.
FT <td>DOMAIN</td> <td>126</td> <td>203</td> <td>IG-LIKE C2-TYPE 1.</td>	DOMAIN	126	203	IG-LIKE C2-TYPE 1.
FT <td>DOMAIN</td> <td>204</td> <td>317</td> <td>IG-LIKE C2-TYPE 2.</td>	DOMAIN	204	317	IG-LIKE C2-TYPE 2.
FT <td>DOMAIN</td> <td>318</td> <td>374</td> <td>IG-LIKE C2-TYPE 3.</td>	DOMAIN	318	374	IG-LIKE C2-TYPE 3.
FT <td>CARBOHYD</td> <td>42</td> <td>42</td> <td>(POTENTIAL).</td>	CARBOHYD	42	42	(POTENTIAL).
FT <td>CARBOHYD</td> <td>286</td> <td>286</td> <td>(BY SIMILARITY).</td>	CARBOHYD	286	286	(BY SIMILARITY).
FT <td>CARBOHYD</td> <td>325</td> <td>325</td> <td>(BY SIMILARITY).</td>	CARBOHYD	325	325	(BY SIMILARITY).
FT <td>DISULFID</td> <td>41</td> <td>109</td> <td>(BY SIMILARITY).</td>	DISULFID	41	109	(BY SIMILARITY).
FT <td>DISULFID</td> <td>155</td> <td>184</td> <td>(BY SIMILARITY).</td>	DISULFID	155	184	(BY SIMILARITY).
FT <td>DISULFID</td> <td>328</td> <td>370</td> <td>(BY SIMILARITY).</td>	DISULFID	328	370	(BY SIMILARITY).
FT <td>LIPID</td> <td>419</td> <td>419</td> <td>(BY SIMILARITY).</td>	LIPID	419	419	(BY SIMILARITY).
FT <td>LIPID</td> <td>422</td> <td>422</td> <td>(BY SIMILARITY).</td>	LIPID	422	422	(BY SIMILARITY).
SO <td>SEQUENCE</td> <td>458 AA.</td> <td>50828 MW;</td> <td>7683EF7F08185535 CR664;</td>	SEQUENCE	458 AA.	50828 MW;	7683EF7F08185535 CR664;
Query Match 54.3%; Score 1853; DB 1; Length 458;				
Best Local Similarity 90.2%; Pred. No. 1.3e-111;				
Matches 357; Conservative 18; Mismatches 21; Indels 0; Gaps 0				
QY	1	MNRGVPFPHLLVLTALIPATOGNKVVLGKKDVTVELCTASQKKSIOFHMKNSNOIK	60	
DB	1	MNRGIPFPHLLVLTALIPATOGKRVVLGKKDVTVELCTASQKKSIOFHMKNSNOIK	60	
QY	61	ILGNQSFLLTKPSKLTLDNRADSRSLMDGCPFLIKNLKIRDSPTVCEVDQKEEYVL	120	
DB	61	ILGIGSFPYTKPSKLTSDPADSRKSLMDGCPFLIKNLKIRDSPTVCEVDQKEEYVL	120	
QY	121	LVFGTLNADPTLLLOGOSLTLTLFLESPPGSSPVOCSRPRGKNIQGGKTLVSQLELDQDSG	180	
DB	121	LVFGTLNADPTLLLOGOSLTLTLFLESPPGSSPVOCSRPRGKNIQGGKTLVSQLELDQDSG	180	
QY	181	TWTCTVLQNOKKVEKIDIVLAFQKASSIVYKKSGEYFSPFLATVEKLTGSGELMW	240	
DB	181	TWTCTVSDQDKVEKIDIVLAFQKASSIVYKKSGEYFSPFLATVEKLTGSGELMW	240	
QY	241	QAEBSASSKSWTTPDLKNKEVSVKRVTOQPKQMGKLPILHTLTPQALQVYAGSGNLTLA	300	
DB	241	QAEBSASSKSWTTPDLKNKEVSVKRVTOQPKQMGKLPILHTLTPQALQVYAGSGNLTLA	300	
QY	301	LEAKTGKLAHOENLVVMRATQLOKQLTCEVMGPTSPKLMLSLKEAKVSKREKPVVY	360	
DB	301	LEAKTGKLAHOENLVVMRATQLOKQLTCEVMGPTSPKLMLSLKEAKVSKREKPVVY	360	
QY	361	LNPEAGMWQCLLSDSGQVLLBSNINVLPTWSTPVED	396	
DB	361	LNPEAGMWQCLLSDSGQVLLBSNINVLPTWSTPVED	396	
RESULT 4				
ID	CD4_MACMU	STANDARD;	PRT;	458 AA.
AC	P16003; Q29617;			
DT	01-APR-1990 (Rel. 14, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	T-cell surface glycoprotein CD4 precursor (T-cell surface antigen			
DE	T4/Leu-3).			
GN	CD4.			
OS	Macaca mulatta (rhesus macaque).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;			
OC	Cercopithecoidea; Macaca.			
OX	NCHI_TaxID=9544;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90182664; PubMed=2107024;			
RA	Camerini D., Seed B.;			
RT	"A CD4 domain important for HIV-mediated syncytium formation lies			
RT	outside the virus binding site.";			
RL	Cell 60:747-754(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Thymocytes;			
RA	Haehimoto O., Tatsumi M.;			

RT	"Molecular cloning and expression of macaque CD4e.";
RL	Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RN	[3]
RP	SEQUENCE OF 28-424 FROM N.A.
RC	TISSUE=Blood;
RX	MEDLINE=93049640; PubMed=1425921;
RA	Pomgaard A., Hirsch V.M., Johnson P.R.;
RT	"Cloning and sequences of primate CD4 molecules: diversity of the
RT	cellular receptor for simian immunodeficiency virus/human
RT	immunodeficiency virus.";
RL	Eur. J. Immunol. 22:2973-2981(1992).
RN	[4]
RP	SEQUENCE OF 107-192 FROM N.A.
RX	MEDLINE=98320644; PubMed=9656488;
RA	Harris E.E., Disotell T.R.;
RT	"Nuclear gene trees and the phylogenetic relationships of the
RT	mangabays (Primates: Papionini).";
RL	Mol. Biol. Evol. 15:892-900(1998).
CC	-I- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC	receptor interaction. May regulate T-cell activation.
CC	-I- SUBUNIT: Associates with p56-1ck (By similarity).
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-I- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC	-I- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL: M31134; AAA36838.1; -
DR	EMBL: D63347; BAA09671.1; -
DR	EMBL: X73326; CAA51752.1; -
DR	EMBL: AF057365; AAC25129.1; -
DR	HSSP: P01730; IMBR.
DR	GO: GO:0042101; C:T-cell receptor complex; ISS.
DR	GO: GO:0015026; F:coreceptor activity; ISS.
DR	GO: GO:0042289; F:MHC class II protein binding; ISS.
DR	GO: GO:0006259; F:Immune response; ISS.
DR	GO: GO:0045086; P:Positive regulation of interleukin-2 biosyn. . .; ISS.
DR	GO: GO:0030217; P:T-cell differentiation; ISS.
DR	GO: GO:0045058; P:T-cell selection; ISS.
DR	GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR	InterPro: IPR000973; CD4 TCRg.
DR	InterPro: IPR007110; Ig_Like.
DR	InterPro: IPR003596; Ig_V.
DR	Pfam: PF00047; Ig; 2.
DR	PRINTS: PR006592; CD4TCANTIGEN.
DR	SMART: SM00406; IGV; 1.
DR	PROSITE: PS50835; Ig_Like; 1.
KW	Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW	Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT	SIGNAL 1 25
FT	CHAIN 26 458
FT	DOMAIN 26 396
FT	TRANSMEM 397 418
FT	DOMAIN 419 458
FT	DOMAIN 126 125
FT	DOMAIN 204 203
FT	DOMAIN 317 317
FT	DOMAIN 374 374
FT	CARBOHYD 296 296
FT	CARBOHYD 325 325
FT	DISULFID 41 109
FT	DISULFID 155 184
FT	DISULFID 328 370
FT	LIPID 419 419
FT	LIPID 422 422
FT	CONFLICT 42 42
FT	CONFLICT 62 62
FT	L -> S (IN REF. 3).



```

CC CONFLICT 67 67 L -> S (IN REF. 2).
CC CONFLICT 169 169 I -> L (IN REF. 2).
CC CONFLICT 191 191 K -> N (IN REF. 3).
CC CONFLICT 248 248 S -> P (IN REF. 2).
CC CONFLICT 265 265 R -> Q (IN REF. 3).
CC CONFLICT 349 349 A -> T (IN REF. 2).
CC CONFLICT 458 AA; 50884 MW; 88880339FAFE808 CRC64;

Query Match
Best Local Similarity 90.2%; Pred. No. 1.5e-11;
Matches 357; Conservative 18; Mismatches 21; Indels 0; Gaps 0;

54.2%; Score 1852; DB 1; Length 458;

OY 1 MNRGVPFRHLLVQLALPPATQGNKVVLGKKGDVTELTCTASQKKSIOFHMKNQIK 60
DB 1 MNRGVPFRHLLVQLALPPATQGNKVVLGKKGDVTELTCTASQKKSIOFHMKNQIK 60
OY 61 ILNGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
DB 61 ILNGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLIEDSDTYICEVENKKEVEL 120
OY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLSVQLELDSG 180
DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLSVQLELDSG 180
OY 181 TWTCVTLOQKKVEFKIDIVLAFQKASSIVYKKEGVEFSPPLATFVEKLTGSGELMW 240
DB 181 TWTCVTLOQKKVEFKIDIVLAFQKASSIVYKKEGVEFSPPLATFVEKLTGSGELMW 240
OY 241 QABRASSSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPHLTLTPQALPYAGSGNLTLLA 300
DB 241 QABRASSSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPHLTLTPQALPYAGSGNLTLLA 300
OY 301 LEAKTGKHOEVNLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKAVSKREKPYVW 360
DB 301 LEAKTGKHOEVNLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKAVSKREKPYVW 360
OY 361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPEP 396
DB 361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPEP 396

RESULT 5
ID CD4_MACFA STANDARD; PRT; 458 AA.
AC P79185;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Leu-3).
GN CD4.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OC NCBI_TaxID=9541;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Thymocytes;
RC Tarsus M., Yabe M., Yamada Y.K.;
RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D63349; BAA09673.1; -.
CC HSSP: P01730; IWR.
DR GO: GO:0042101; C: T-cell receptor complex; ISS.
DR GO: GO:0015026; F: coreceptor activity; ISS.
DR GO: GO:0042289; F: MHC class II protein binding; ISS.
DR GO: GO:0006955; P: immune response; ISS.
DR GO: GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO: GO:0030217; P: T-cell differentiation; ISS.
DR GO: GO:0045058; P: T-cell selection; ISS.
DR GO: GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4_TCAg.
DR InterPro: IPR007110; Ig_Like.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IG_v_1.
DR PROSITE: PS50835; IG_LIKE_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 42 42
FT CARBOHYD 296 296
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
SQ SEQUENCE 458 AA; 50872 MW; 9105479FB556FE7 CRC64;

Query Match
Best Local Similarity 90.2%; Pred. No. 2.3e-11;
Matches 357; Conservative 17; Mismatches 22; Indels 0; Gaps 0;

54.2%; Score 1849; DB 1; Length 458;

OY 1 MNRGVPFRHLLVQLALPPATQGNKVVLGKKGDVTELTCTASQKKSIOFHMKNQIK 60
DB 1 MNRGVPFRHLLVQLALPPATQGNKVVLGKKGDVTELTCTASQKKSIOFHMKNQIK 60
OY 61 ILNGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120
DB 61 ILNGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLIEDSDTYICEVENKKEVEL 120
OY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLSVQLELDSG 180
DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLSVQLELDSG 180
OY 181 TWTCVTLOQKKVEFKIDIVLAFQKASSIVYKKEGVEFSPPLATFVEKLTGSGELMW 240
DB 181 TWTCVTLOQKKVEFKIDIVLAFQKASSIVYKKEGVEFSPPLATFVEKLTGSGELMW 240
OY 241 QABRASSSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPHLTLTPQALPYAGSGNLTLLA 300
DB 241 QABRASSSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPHLTLTPQALPYAGSGNLTLLA 300
OY 301 LEAKTGKHOEVNLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKAVSKREKPYVW 360
DB 301 LEAKTGKHOEVNLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKAVSKREKPYVW 360
OY 361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPEP 396
DB 361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPEP 396

```



```

RESULT 6
CD4_MACNE
ID CD4_MACNE STANDARD; PRT; 458 AA.
AC 008340; P79196;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Leu-3).
GN CD4.
OS Macaca nemestrina (Pig-tailed macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OC NCBI_TaxId=9545;
RX [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatsumi M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RX Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus."
RL Eur. J. Immunol. 22:2973-2981 (1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D63346; BA09670.1; -
DR EMBL: X73325; CA51751.1; -
DR HSSP: P01730; IWR.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; P:coreceptor activity; ISS.
DR GO: GO:0042289; P:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.
DR InterPro: IPR000973; CD4_Tcag.
DR InterPro: IPR007110; Ig_Like.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV_1.
DR PROSITE: PS50835; IGV_Like_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT CHAIN 1
FT SIGNAL 25
FT CHAIN 26
FT DOMAIN 26
FT TRANSMEM 397
FT DOMAIN 419
FT DOMAIN 426
FT DOMAIN 126
FT DOMAIN 204
FT DOMAIN 318
FT DOMAIN 374
IG-LIKE C2-TYPE 1.
IG-LIKE C2-TYPE 2.
IG-LIKE C2-TYPE 3.

```

```

FT CARBOHYD 42
FT CARBOHYD 296
FT CARBOHYD 325
FT DISULFID 41
FT DISULFID 155
FT DISULFID 328
FT DISULFID 419
FT LIPID 422
FT LIPID 57
FT LIPID 91
FT LIPID 105
FT LIPID 113
FT LIPID 302
FT LIPID 349
FT LIPID 395
FT LIPID 422
FT LIPID 449
FT LIPID 475
FT LIPID 501
FT LIPID 527
FT LIPID 553
FT LIPID 579
FT LIPID 605
FT LIPID 631
FT LIPID 657
FT LIPID 683
FT LIPID 709
FT LIPID 735
FT LIPID 761
FT LIPID 787
FT LIPID 813
FT LIPID 839
FT LIPID 865
FT LIPID 891
FT LIPID 917
FT LIPID 943
FT LIPID 969
FT LIPID 995
FT LIPID 1021
FT LIPID 1047
FT LIPID 1073
FT LIPID 1099
FT LIPID 1125
FT LIPID 1151
FT LIPID 1177
FT LIPID 1203
FT LIPID 1229
FT LIPID 1255
FT LIPID 1281
FT LIPID 1307
FT LIPID 1333
FT LIPID 1359
FT LIPID 1385
FT LIPID 1411
FT LIPID 1437
FT LIPID 1463
FT LIPID 1489
FT LIPID 1515
FT LIPID 1541
FT LIPID 1567
FT LIPID 1593
FT LIPID 1619
FT LIPID 1645
FT LIPID 1671
FT LIPID 1697
FT LIPID 1723
FT LIPID 1749
FT LIPID 1775
FT LIPID 1801
FT LIPID 1827
FT LIPID 1853
FT LIPID 1879
FT LIPID 1905
FT LIPID 1931
FT LIPID 1957
FT LIPID 1983
FT LIPID 2009
FT LIPID 2035
FT LIPID 2061
FT LIPID 2087
FT LIPID 2113
FT LIPID 2139
FT LIPID 2165
FT LIPID 2191
FT LIPID 2217
FT LIPID 2243
FT LIPID 2269
FT LIPID 2295
FT LIPID 2321
FT LIPID 2347
FT LIPID 2373
FT LIPID 2399
FT LIPID 2425
FT LIPID 2451
FT LIPID 2477
FT LIPID 2503
FT LIPID 2529
FT LIPID 2555
FT LIPID 2581
FT LIPID 2607
FT LIPID 2633
FT LIPID 2659
FT LIPID 2685
FT LIPID 2711
FT LIPID 2737
FT LIPID 2763
FT LIPID 2789
FT LIPID 2815
FT LIPID 2841
FT LIPID 2867
FT LIPID 2893
FT LIPID 2919
FT LIPID 2945
FT LIPID 2971
FT LIPID 2997
FT LIPID 3023
FT LIPID 3049
FT LIPID 3075
FT LIPID 3101
FT LIPID 3127
FT LIPID 3153
FT LIPID 3179
FT LIPID 3205
FT LIPID 3231
FT LIPID 3257
FT LIPID 3283
FT LIPID 3309
FT LIPID 3335
FT LIPID 3361
FT LIPID 3387
FT LIPID 3413
FT LIPID 3439
FT LIPID 3465
FT LIPID 3491
FT LIPID 3517
FT LIPID 3543
FT LIPID 3569
FT LIPID 3595
FT LIPID 3621
FT LIPID 3647
FT LIPID 3673
FT LIPID 3699
FT LIPID 3725
FT LIPID 3751
FT LIPID 3777
FT LIPID 3803
FT LIPID 3829
FT LIPID 3855
FT LIPID 3881
FT LIPID 3907
FT LIPID 3933
FT LIPID 3959
FT LIPID 3985
FT LIPID 4011
FT LIPID 4037
FT LIPID 4063
FT LIPID 4089
FT LIPID 4115
FT LIPID 4141
FT LIPID 4167
FT LIPID 4193
FT LIPID 4219
FT LIPID 4245
FT LIPID 4271
FT LIPID 4297
FT LIPID 4323
FT LIPID 4349
FT LIPID 4375
FT LIPID 4401
FT LIPID 4427
FT LIPID 4453
FT LIPID 4479
FT LIPID 4505
FT LIPID 4531
FT LIPID 4557
FT LIPID 4583
FT LIPID 4609
FT LIPID 4635
FT LIPID 4661
FT LIPID 4687
FT LIPID 4713
FT LIPID 4739
FT LIPID 4765
FT LIPID 4791
FT LIPID 4817
FT LIPID 4843
FT LIPID 4869
FT LIPID 4895
FT LIPID 4921
FT LIPID 4947
FT LIPID 4973
FT LIPID 4999
FT LIPID 5025
FT LIPID 5051
FT LIPID 5077
FT LIPID 5103
FT LIPID 5129
FT LIPID 5155
FT LIPID 5181
FT LIPID 5207
FT LIPID 5233
FT LIPID 5259
FT LIPID 5285
FT LIPID 5311
FT LIPID 5337
FT LIPID 5363
FT LIPID 5389
FT LIPID 5415
FT LIPID 5441
FT LIPID 5467
FT LIPID 5493
FT LIPID 5519
FT LIPID 5545
FT LIPID 5571
FT LIPID 5597
FT LIPID 5623
FT LIPID 5649
FT LIPID 5675
FT LIPID 5701
FT LIPID 5727
FT LIPID 5753
FT LIPID 5779
FT LIPID 5805
FT LIPID 5831
FT LIPID 5857
FT LIPID 5883
FT LIPID 5909
FT LIPID 5935
FT LIPID 5961
FT LIPID 5987
FT LIPID 6013
FT LIPID 6039
FT LIPID 6065
FT LIPID 6091
FT LIPID 6117
FT LIPID 6143
FT LIPID 6169
FT LIPID 6195
FT LIPID 6221
FT LIPID 6247
FT LIPID 6273
FT LIPID 6299
FT LIPID 6325
FT LIPID 6351
FT LIPID 6377
FT LIPID 6403
FT LIPID 6429
FT LIPID 6455
FT LIPID 6481
FT LIPID 6507
FT LIPID 6533
FT LIPID 6559
FT LIPID 6585
FT LIPID 6611
FT LIPID 6637
FT LIPID 6663
FT LIPID 6689
FT LIPID 6715
FT LIPID 6741
FT LIPID 6767
FT LIPID 6793
FT LIPID 6819
FT LIPID 6845
FT LIPID 6871
FT LIPID 6897
FT LIPID 6923
FT LIPID 6949
FT LIPID 6975
FT LIPID 7001
FT LIPID 7027
FT LIPID 7053
FT LIPID 7079
FT LIPID 7105
FT LIPID 7131
FT LIPID 7157
FT LIPID 7183
FT LIPID 7209
FT LIPID 7235
FT LIPID 7261
FT LIPID 7287
FT LIPID 7313
FT LIPID 7339
FT LIPID 7365
FT LIPID 7391
FT LIPID 7417
FT LIPID 7443
FT LIPID 7469
FT LIPID 7495
FT LIPID 7521
FT LIPID 7547
FT LIPID 7573
FT LIPID 7599
FT LIPID 7625
FT LIPID 7651
FT LIPID 7677
FT LIPID 7703
FT LIPID 7729
FT LIPID 7755
FT LIPID 7781
FT LIPID 7807
FT LIPID 7833
FT LIPID 7859
FT LIPID 7885
FT LIPID 7911
FT LIPID 7937
FT LIPID 7963
FT LIPID 7989
FT LIPID 8015
FT LIPID 8041
FT LIPID 8067
FT LIPID 8093
FT LIPID 8119
FT LIPID 8145
FT LIPID 8171
FT LIPID 8197
FT LIPID 8223
FT LIPID 8249
FT LIPID 8275
FT LIPID 8301
FT LIPID 8327
FT LIPID 8353
FT LIPID 8379
FT LIPID 8405
FT LIPID 8431
FT LIPID 8457
FT LIPID 8483
FT LIPID 8509
FT LIPID 8535
FT LIPID 8561
FT LIPID 8587
FT LIPID 8613
FT LIPID 8639
FT LIPID 8665
FT LIPID 8691
FT LIPID 8717
FT LIPID 8743
FT LIPID 8769
FT LIPID 8795
FT LIPID 8821
FT LIPID 8847
FT LIPID 8873
FT LIPID 8899
FT LIPID 8925
FT LIPID 8951
FT LIPID 8977
FT LIPID 9003
FT LIPID 9029
FT LIPID 9055
FT LIPID 9081
FT LIPID 9107
FT LIPID 9133
FT LIPID 9159
FT LIPID 9185
FT LIPID 9211
FT LIPID 9237
FT LIPID 9263
FT LIPID 9289
FT LIPID 9315
FT LIPID 9341
FT LIPID 9367
FT LIPID 9393
FT LIPID 9419
FT LIPID 9445
FT LIPID 9471
FT LIPID 9497
FT LIPID 9523
FT LIPID 9549
FT LIPID 9575
FT LIPID 9601
FT LIPID 9627
FT LIPID 9653
FT LIPID 9679
FT LIPID 9705
FT LIPID 9731
FT LIPID 9757
FT LIPID 9783
FT LIPID 9809
FT LIPID 9835
FT LIPID 9861
FT LIPID 9887
FT LIPID 9913
FT LIPID 9939
FT LIPID 9965
FT LIPID 9991
FT LIPID 10007
FT LIPID 10033
FT LIPID 10059
FT LIPID 10085
FT LIPID 10111
FT LIPID 10137
FT LIPID 10163
FT LIPID 10189
FT LIPID 10215
FT LIPID 10241
FT LIPID 10267
FT LIPID 10293
FT LIPID 10319
FT LIPID 10345
FT LIPID 10371
FT LIPID 10397
FT LIPID 10423
FT LIPID 10449
FT LIPID 10475
FT LIPID 10501
FT LIPID 10527
FT LIPID 10553
FT LIPID 10579
FT LIPID 10605
FT LIPID 10631
FT LIPID 10657
FT LIPID 10683
FT LIPID 10709
FT LIPID 10735
FT LIPID 10761
FT LIPID 10787
FT LIPID 10813
FT LIPID 10839
FT LIPID 10865
FT LIPID 10891
FT LIPID 10917
FT LIPID 10943
FT LIPID 10969
FT LIPID 10995
FT LIPID 11021
FT LIPID 11047
FT LIPID 11073
FT LIPID 11099
FT LIPID 11125
FT LIPID 11151
FT LIPID 11177
FT LIPID 11203
FT LIPID 11229
FT LIPID 11255
FT LIPID 11281
FT LIPID 11307
FT LIPID 11333
FT LIPID 11359
FT LIPID 11385
FT LIPID 11411
FT LIPID 11437
FT LIPID 11463
FT LIPID 11489
FT LIPID 11515
FT LIPID 11541
FT LIPID 11567
FT LIPID 11593
FT LIPID 11619
FT LIPID 11645
FT LIPID 11671
FT LIPID 11697
FT LIPID 11723
FT LIPID 11749
FT LIPID 11775
FT LIPID 11801
FT LIPID 11827
FT LIPID 11853
FT LIPID 11879
FT LIPID 11905
FT LIPID 11931
FT LIPID 11957
FT LIPID 11983
FT LIPID 12009
FT LIPID 12035
FT LIPID 12061
FT LIPID 12087
FT LIPID 12113
FT LIPID 12139
FT LIPID 12165
FT LIPID 12191
FT LIPID 12217
FT LIPID 12243
FT LIPID 12269
FT LIPID 12295
FT LIPID 12321
FT LIPID 12347
FT LIPID 12373
FT LIPID 12399
FT LIPID 12425
FT LIPID 12451
FT LIPID 12477
FT LIPID 12503
FT LIPID 12529
FT LIPID 12555
FT LIPID 12581
FT LIPID 12607
FT LIPID 12633
FT LIPID 12659
FT LIPID 12685
FT LIPID 12711
FT LIPID 12737
FT LIPID 12763
FT LIPID 12789
FT LIPID 12815
FT LIPID 12841
FT LIPID 12867
FT LIPID 12893
FT LIPID 12919
FT LIPID 12945
FT LIPID 12971
FT LIPID 12997
FT LIPID 13023
FT LIPID 13049
FT LIPID 13075
FT LIPID 13101
FT LIPID 13127
FT LIPID 13153
FT LIPID 13179
FT LIPID 13205
FT LIPID 13231
FT LIPID 13257
FT LIPID 13283
FT LIPID 13309
FT LIPID 13335
FT LIPID 13361
FT LIPID 13387
FT LIPID 13413
FT LIPID 13439
FT LIPID 13465
FT LIPID 13491
FT LIPID 13517
FT LIPID 13543
FT LIPID 13569
FT LIPID 13595
FT LIPID 13621
FT LIPID 13647
FT LIPID 13673
FT LIPID 13699
FT LIPID 13725
FT LIPID 13751
FT LIPID 13777
FT LIPID 13803
FT LIPID 13829
FT LIPID 13855
FT LIPID 13881
FT LIPID 13907
FT LIPID 13933
FT LIPID 13959
FT LIPID 13985
FT LIPID 14011
FT LIPID 14037
FT LIPID 14063
FT LIPID 14089
FT LIPID 14115
FT LIPID 14141
FT LIPID 14167
FT LIPID 14193
FT LIPID 14219
FT LIPID 14245
FT LIPID 14271
FT LIPID 14297
FT LIPID 14323
FT LIPID 14349
FT LIPID 14375
FT LIPID 14401
FT LIPID 14427
FT LIPID 14453
FT LIPID 14479
FT LIPID 14505
FT LIPID 14531
FT LIPID 14557
FT LIPID 14583
FT LIPID 14609
FT LIPID 14635
FT LIPID 14661
FT LIPID 14687
FT LIPID 14713
FT LIPID 14739
FT LIPID 14765
FT LIPID 14791
FT LIPID 14817
FT LIPID 14843
FT LIPID 14869
FT LIPID 14895
FT LIPID 14921
FT LIPID 14947
FT LIPID 14973
FT LIPID 14999
FT LIPID 15025
FT LIPID 15051
FT LIPID 15077
FT LIPID 15103
FT LIPID 15129
FT LIPID 15155
FT LIPID 15181
FT LIPID 15207
FT LIPID 15233
FT LIPID 15259
FT LIPID 15285
FT LIPID 15311
FT LIPID 15337
FT LIPID 15363
FT LIPID 15389
FT LIPID 15415
FT LIPID 15441
FT LIPID 15467
FT LIPID 15493
FT LIPID 15519
FT LIPID 15545
FT LIPID 15571
FT LIPID 15597
FT LIPID 15623
FT LIPID 15649
FT LIPID 15675
FT LIPID 15701
FT LIPID 15727
FT LIPID 15753
FT LIPID 15779
FT LIPID 15805
FT LIPID 15831
FT LIPID 15857
FT LIPID 15883
FT LIPID 15909
FT LIPID 15935
FT LIPID 15961
FT LIPID 15987
FT LIPID 16013
FT LIPID 16039
FT LIPID 16065
FT LIPID 16091
FT LIPID 16117
FT LIPID 16143
FT LIPID 16169
FT LIPID 16195
FT LIPID 16221
FT LIPID 16247
FT LIPID 16273
FT LIPID 16299
FT LIPID 16325
FT LIPID 16351
FT LIPID 16377
FT LIPID 16403
FT LIPID 16429
FT LIPID 16455
FT LIPID 16481
FT LIPID 16507
FT LIPID 16533
FT LIPID 16559
FT LIPID 16585
FT LIPID 16611
FT LIPID 16637
FT LIPID 16663
FT LIPID 16689
FT LIPID 16715
FT LIPID 16741
FT LIPID 16767
FT LIPID 16793
FT LIPID 16819
FT LIPID 16845
FT LIPID 16871
FT LIPID 16897
FT LIPID 16923
FT LIPID 16949
FT LIPID 16975
FT LIPID 16999
FT LIPID 17025
FT LIPID 17051
FT LIPID 17077
FT LIPID 17103
FT LIPID 17129
FT LIPID 17155
FT LIPID 17181
FT LIPID 17207
FT LIPID 17233
FT LIPID 17259
FT LIPID 17285
FT LIPID 17311
FT LIPID 17337
FT LIPID 17363
FT LIPID 17389
FT LIPID 17415
FT LIPID 17441
FT LIPID 17467
FT LIPID 17493
FT LIPID 17519
FT LIPID 17545
FT LIPID 17571
FT LIPID 17597
FT LIPID 17623
FT LIPID 17649
FT LIPID 17675
FT LIPID 17701
FT LIPID 17727
FT LIPID 17753
FT LIPID 17779
FT LIPID 17805
FT LIPID 17831
FT LIPID 17857
FT LIPID 17883
FT LIPID 17909
FT LIPID 17935
FT LIPID 17961
FT LIPID 17987
FT LIPID 18013
FT LIPID 18039
FT LIPID 18065
FT LIPID 18091
FT LIPID 18117
FT LIPID 18143
FT LIPID 18169
FT LIPID 18195
FT LIPID 18221
FT LIPID 18247
FT LIPID 18273
FT LIPID 18299
FT LIPID 18325
FT LIPID 18351
FT LIPID 18377
FT LIPID 18403
FT LIPID 18429
FT LIPID 18455
FT LIPID 18481
FT LIPID 18507
FT LIPID 18533
FT LIPID 18559
FT LIPID 18585
FT LIPID 18611
FT LIPID 18637
FT LIPID 18663
FT LIPID 18689
FT LIPID 18715
FT LIPID 18741
FT LIPID 18767
FT LIPID 18793
FT LIPID 18819
FT LIPID 18845
FT LIPID 18871
FT LIPID 18897
FT LIPID 18923
FT LIPID 18949
FT LIPID 18975
FT LIPID 18999
FT LIPID 19025
FT LIPID 19051
FT LIPID 19077
FT LIPID 19103
FT LIPID 19129
FT LIPID 19155
FT LIPID 19181
FT LIPID 19207
FT LIPID 19233
FT LIPID 19259
FT LIPID 19285
FT LIPID 19311
FT LIPID 19337
FT LIPID 19363
FT LIPID 19389
FT LIPID 19415
FT LIPID 19441
FT LIPID 19467
FT LIPID 19493
FT LIPID 19519
FT LIPID 19545
FT LIPID 19571
FT LIPID 19597
FT LIPID 19623
FT LIPID 19649
FT LIPID 19675
FT LIPID 19701
FT LIPID 19727
FT LIPID 19753
FT LIPID 19779
FT LIPID 19805
FT LIPID 19831
FT LIPID 19857
FT LIPID 19883
FT LIPID 19909
FT LIPID 19935
FT LIPID 19961
FT LIPID 19987
FT LIPID 20013
FT LIPID 20039
FT LIPID 20065
FT LIPID 20091
FT LIPID 20117
FT LIPID 20143
FT LIPID 20169
FT LIPID 20195
FT LIPID 20221
FT LIPID 20247
FT LIPID 20273
FT LIPID 20299
FT LIPID 20325
FT LIPID 20351
FT LIPID 20377
FT LIPID 20403
FT LIPID 20429
FT LIPID 20455
FT LIPID 20481
FT LIPID 20507
FT LIPID 20533
FT LIPID 20559
FT LIPID 20585
FT LIPID 20611
FT LIPID 20637
FT LIPID 20663
FT LIPID 20689
FT LIPID 20715
FT LIPID 20741
FT LIPID 20767
FT LIPID 20793
FT LIPID 20819
FT LIPID 20845
FT LIPID 20871
FT LIPID 20897
FT LIPID 20923
FT LIPID 20949
FT LIPID 20975
FT LIPID 20999
FT LIPID 21025
FT LIPID 21051
FT LIPID 21077
FT LIPID 21103
FT LIPID 21129
FT LIPID 21155
FT LIPID 21181
FT LIPID 21207
FT LIPID 21233
FT LIPID 21259
FT LIPID 21285
FT LIPID 21311
FT LIPID 21337
FT LIPID 21363
FT LIPID 21389
FT LIPID 21415
FT LIPID 21441
FT LIPID 21467
FT LIPID 21493
FT LIPID 21519
FT LIPID 21545
FT LIPID 21571
FT LIPID 21597
FT LIPID 21623
FT LIPID 21649
FT LIPID 21675
FT LIPID 21701
FT LIPID 21727
FT LIPID 21753
FT LIPID 21779
FT LIPID 21805
FT LIPID 21831
FT LIPID 21857
FT LIPID 21883
FT LIPID 21909
FT LIPID 21935
FT LIPID 21961
FT LIPID 21987
FT LIPID 22013
FT LIPID 22039
FT LIPID 22065
FT LIPID 22091
FT LIPID 22117
FT LIPID 22143
FT LIPID 22169
FT LIPID 22195
FT LIPID 22221
FT LIPID 22247
FT LIPID 22273
FT LIPID 22299
FT LIPID 22325
FT LIPID 22351
FT LIPID 22377
FT LIPID 22403
FT LIPID 22429
FT LIPID 22455
FT LIPID 22481
FT LIPID 22507
FT LIPID 22533
FT LIPID 22559
FT LIPID 22585
FT LIPID 22611
FT LIPID 22637
FT LIPID 22663
FT LIPID 22689
FT LIPID 22715
FT LIPID 22741
FT LIPID 22767
FT LIPID 22793
FT LIPID 22819
FT LIPID 22845
FT LIPID 22871
FT LIPID 22897
FT LIPID 22923
FT LIPID 22949
FT LIPID 22975
FT LIPID 22999
FT LIPID 23025
FT LIPID 23051
FT LIPID 23077
FT LIPID 23103
FT LIPID 23129
FT LIPID 23155
FT LIPID 23181
FT LIPID 23207
FT LIPID 23233
FT LIPID 23259
FT LIPID 23285
FT LIPID 23311
FT LIPID 23337
FT LIPID 23363
FT LIPID 23389
FT LIPID 23415
FT LIPID 23441
FT LIPID 23467
FT LIPID 23493
FT LIPID 23519
FT LIPID 23545
FT LIPID 23571
FT LIPID 23597
FT LIPID 23623
FT LIPID 23649
FT LIPID 23675
FT LIPID 23701
FT LIPID 23727
FT LIPID 23753
FT LIPID 23779
FT LIPID 23805
FT LIPID 23831
FT LIPID 23857
FT LIPID 23883
FT LIPID 23909
FT LIPID 23935
FT LIPID 23961
FT LIPID 23987
FT LIPID 24013
FT LIPID 24039
FT LIPID 24065
FT LIPID 24091
FT LIPID 24117
FT LIPID 24143
FT LIPID 24169
FT LIPID 24195
FT LIPID 24221
FT LIPID 24247
FT LIPID 24273
FT LIPID 24299
FT LIPID 24325
FT LIPID 24351
FT LIPID 24377
FT LIPID 24403
FT LIPID 24429
FT LIPID 24455
FT LIPID 24481
FT LIPID 24507
FT LIPID 24533
FT LIPID 24559
FT LIPID 24585
FT LIPID 24611
FT LIPID 24637
FT LIPID 24663
FT LIPID 24689
FT LIPID 24715
FT LIPID 24741
FT LIPID 24767
FT LIPID 24793
FT LIPID 24819
FT LIPID 24845
FT LIPID 24871
FT LIPID 24897
FT LIPID 24923
FT LIPID 24949
FT LIPID 24975
FT LIPID 24999
FT LIPID 25025
FT LIPID 25051
FT LIPID 25077
FT LIPID 25103
FT LIPID 25129
FT LIPID 25155
FT LIPID 25181
FT LIPID 25207
FT LIPID 25233
FT LIPID 25259
FT LIPID 25285
FT LIPID 25311
FT LIPID 25337
FT LIPID 25363
FT LIPID 25389
FT LIPID 25415
FT LIPID 25441
FT LIPID 25467
FT LIPID 25493
FT LIPID 25519
FT LIPID 25545
FT LIPID 25571
FT LIPID 25597
FT LIPID 25623
FT LIPID 25649
FT LIPID 25675
FT LIPID 25701
FT LIPID 25727
FT LIPID 25753
FT LIPID 25779
FT LIPID 25805
FT LIPID 25831
FT LIPID 25857
FT LIPID 25883
FT LIPID 25909
FT LIPID 25935
FT LIPID 25961
FT LIPID 25987
FT LIPID 26013
FT LIPID 26039
FT LIPID 26065
FT LIPID 26091
FT LIPID 26117
FT LIPID 26143
FT LIPID 26169
FT LIPID 26195
FT LIPID 26221
FT LIPID 26247
FT LIPID 26273
FT LIPID 26299
FT LIPID 26325
FT LIPID 26351
FT LIPID 26377
FT LIPID 26403
FT LIPID 26429
FT LIPID 26455
FT LIPID 26481
FT LIPID 26507
FT LIPID 26533
FT LIPID 26559
FT LIPID 26585
FT LIPID 26611
FT LIPID 26637
FT LIPID 26663
FT LIPID 26689
FT LIPID 26715
FT LIPID 26741
FT LIPID 26767
FT LIPID 26793
FT LIPID 26819
FT LIPID 26845
FT LIPID 26871
FT LIPID 26897
FT LIPID 26923
FT LIPID 26949
FT LIPID 26975
FT LIP
```



RA Fomsgaard A., Hirsch V.M., Johnson P.R.;  
 RT "Cloning and sequences of primate CD4 molecules: diversity of the  
 RT cellular receptor for simian immunodeficiency virus/human  
 RT immunodeficiency virus.";  
 RL Eur. J. Immunol. 22:2973-2981(1992).  
 RN [3]  
 RP SEQUENCE OF 28-424 FROM N.A.  
 RC TISSUE=peripheral blood;  
 RX Fomsgaard A., Mueller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,  
 RA Corbet S., Barre-Sinoussi F., Allan J.S.;  
 RT "Relation between phylogeny of African green monkey CD4 genes and  
 RT their respective simian immunodeficiency virus genes.";  
 RL J. Med. Primatol. 26:120-128(1997).  
 RN [4]  
 RP SEQUENCE OF 107-192 FROM N.A.  
 RX MEDLINE=98320644; PubMed=9656488;  
 RA Harris E.E., Disotell T.R.;  
 RT "Nuclear gene trees and the phylogenetic relationships of the  
 RT mangabeys (Primates: Papionini).";  
 RL Mol. Biol. Evol. 15:892-900(1998).  
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell  
 CC receptor interaction. May regulate T-cell activation.  
 CC -1- SUBUNIT: Associates with p56<sup>lck</sup> (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, D86589, BAAL3132.1, -;  
 DR EMBL, X73322, CA51748.1, -;  
 DR EMBL, AF001226, AAB60873.1, -;  
 DR EMBL, AF001228, AAB60875.1, -;  
 DR EMBL, AF057380, AAC25124.1, -;  
 DR HSRP, P01730, IWO.  
 DR GO, GO:0042101, C-T-cell receptor complex; ISS.  
 DR GO, GO:0015026, F:coreceptor activity; ISS.  
 DR GO, GO:0042289, F:MHC class II protein binding; ISS.  
 DR GO, GO:0006955, P:immune response; ISS.  
 DR GO, GO:0045086, P:positive regulation of interleukin-2 biosyn. . . ; ISS.  
 DR GO, GO:0030217, P:T-cell differentiation; ISS.  
 DR GO, GO:0045058, P:T-cell selection; ISS.  
 DR GO, GO:0007169, P:transmembrane receptor protein tyrosine kin. . . ; ISS.  
 DR InterPro: IPR000973; CD4 TCAG.  
 DR InterPro: IPR007110, IG-like.  
 DR InterPro: IPR003596; IG\_v.  
 DR Pfam: PF00047, Ig\_2.  
 DR PRINTS, PRO0692, CD4TCANTISEN.  
 DR SMART, SM00406, IGV, 1.  
 DR PROSITE, PS50835, IG\_LIKE, 1.  
 DR Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;  
 KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.  
 FT SIGNAL 1 25  
 FT CHAIN 26 458  
 FT DOMAIN 26 396  
 FT TRANSMEM 397 418  
 FT DOMAIN 419 458  
 FT DOMAIN 26 125  
 FT DOMAIN 204 317  
 FT DOMAIN 317 374  
 FT CARBOHYD 42 42  
 FT CARBOHYD 281 281  
 FT CARBOHYD 296 296  
 FT CARBOHYD 325 325  
 FT DISULFID 41 109  
 BY SIMILARITY.

FT DISULFID 155 184 BY SIMILARITY.  
 FT DISULFID 328 370 BY SIMILARITY.  
 FT LIPID 419 419 S-palmitoyl cysteine (By similarity).  
 FT LIPID 422 422 S-palmitoyl cysteine (By similarity).  
 FT CONFLICT 46 46 K -> N (IN REF. 2 AND 3; AAB60875).  
 FT CONFLICT 59 59 I -> T (IN REF. 3; AAB60873).  
 FT CONFLICT 115 115 K -> E (IN REF. 1).  
 FT CONFLICT 165 165 G -> V (IN REF. 3; AAB60873 AND 4).  
 FT CONFLICT 200 200 M -> V (IN REF. 2 AND 3).  
 FT CONFLICT 227 227 F -> L (IN REF. 3; AAB60873).  
 FT CONFLICT 271 271 K -> E (IN REF. 3; AAB60873).  
 FT CONFLICT 281 281 N -> H (IN REF. 3; AAB60873).  
 SQ SEQUENCE 458 AA, 51158 MW, FC523D2EDDLF72E7 CRC64;  
 Query Match 53.5%; Score 1827; DB 1; Length 458;  
 Best Local Similarity 89.1%; Pred. No. 5; ee-110;  
 Matches 355; Conservative 18; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 MNRGVPFRHLVLVQLALPAATQGNKRVLGKGDVTELTCTASQKKSIOFHKNSNOIK 60  
 DB 1 MNMGIPFRHLVLVQLALPAVTOGKKVVGKGDVTELTCTASQKKTTFPHWNSNOIK 60  
 QY 61 ILGQGSFLTKGSPKLNDRADRSRLMDGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120  
 DB 61 ILGQGSFLTKGSKLNDRADRSRLMDQCFSTIKNLKIEDSDTYICEVENKKEVEL 120  
 QY 121 LVFGLTANSDTHLLOGSLTLTLSPPGSSPSVQCRSPRKNIOGKTLVSQLELDSG 180  
 DB 121 LVFGLTANSDTHLLOGSLTLTLSPPGSSPSVQCRSPRKNIOGKTLVSQLELDSG 180  
 QY 181 TWTCVTVQNDKQVFEKIDIVLAFQKASSIVYKKEGQVESPPLATVEKLTSGGLMW 240  
 DB 181 TWTCVTVQNDKQVFEKIDIVLAFQKASSIVYKKEGQVESPPLATVEKLTSGGLMW 240  
 QY 241 QABRASSKSMITFDLKNKEVSVKRVTOQPKLQMGKPLPHLTLPOLPOYAGSGNTLA 300  
 DB 241 QABRASSKSMITFDLKNKEVSVKQVTOQPKLQMGKPLPHLTLPOLPOYAGSGNTLA 300  
 QY 301 LEAKTGLHOEVNLVWRATOLQKNLTCEVGEPTSPKMLSTLKNKAAYSKREKPVW 360  
 DB 301 LEAKTGLHOEVNLVWRATOLQKNLTCEVGEPTSPKMLSTLKNKAAYSKREKPVW 360  
 QY 361 LNPEAGWQCLSDSGVLLSNIKVLPTWSTPEP 396  
 DB 361 LNPEAGWQCLSDSGVLLSNIKVLPTWSTPEP 396  
 RESULT 8  
 CD4\_ERYPA STANDARD; PRT; 397 AA.  
 AC Q08339;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3)  
 DE (Fragment).  
 GN CD4.  
 OS Erythrocytus patas (Red quonon) (Huesar).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Erythrocytus.  
 OX NCB1\_TaxID=9538;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Blood;  
 RX MEDLINE=93049640; PubMed=1425921;  
 RA Fomsgaard A., Hirsch V.M., Johnson P.R.;  
 RT "Cloning and sequences of primate CD4 molecules: diversity of the  
 RT cellular receptor for simian immunodeficiency virus/human  
 RT immunodeficiency virus.";  
 RL Eur. J. Immunol. 22:2973-2981(1992).  
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell  
 CC receptor interaction. May regulate T-cell activation.



```

CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X73324; CAAS1750.1; -.
CC HSP: P01730; IWIQ.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:receptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR007110; Ig_Like.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_2.
DR SMART: SMART; CD4TCANTIGEN.
DR SMART: SM00406; IGV; 1.
DR PROSITE: PS00835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Lipoprotein; Palmitate.
FT DOMAIN 1 369
FT TRANSMEM 370 391
FT DOMAIN 392 >397
FT DOMAIN <1 98
FT DOMAIN 99 176
FT DOMAIN 177 290
FT DOMAIN 291 347
FT DOMAIN 269 269
FT CARBOHYD 298 298
FT CARBOHYD 14 82
FT DISULFID 128 157
FT DISULFID 301 343
FT LIPID 392 392
FT LIPID 395 395
FT LIPID 397 397
FT NON_TER 397 397
SQ SEQUENCE 397 AA; 44081 MW; 67887397A6B7EA4F CRC64;

Query Match 50.6%; Score 1726; DB 1; Length 397;
Best Local Similarity 89.7%; Pred. No. 1,4e-103;
Matches 331; Conservative 16; Mismatches 22; Indels 0; Gaps 0;

28 VVLGGKGDVVELTCTASOKKSIQPHKNSNOKITLGNQSPITLKPSKINDRADSRSM 87
1 VVLGGKGDVVELTCTASOKKSIQPHKNSNOKITLGNQSPITLKPSKINDRADSRSM 60
88 DQGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSPTDHLQGSQTLTLESP 147
61 DQGNFPLIKNLKIEDSEITYICEVEDQKEVQLVFGLTANSPTDHLQGSQTLTLESP 120
148 GSSPSVQCSPPKRNKIQGGKTLSSVSLQDSTWTCTYLQNKQKYEKIDIVLAFOQA 207
121 GSSPSVQCSPPKRNKIQGGKTLSSVSLQDSTWTCTYLQNKQKYEKIDIVLAFOQA 180
208 SSIIVYKGEQYEFSPFLAFVTEKLLGSGELMMQARASSSKSWITFDLKNKEVSRYRT 267
181 SSIIVYKGEQYEFSPFLAFVTEKLLGSGELMMQARASSSKSWITFDLKNKEVSRYRT 240
268 QDPKLOMGKKLPLHLTLPLPOLPOLYAGSGNLTALAEKTKLQEVNLVYMRATOLQKNLT 327
241 QDPKLOMGKKLPLHLTLPLPOLPOLYAGSGNLTALAEKTKLQEVNLVYMRATOLQKNLT 300

```

```

QY 328 CEVWGPSPKMLSLKENKEAKVSKREKPVWVINEPAGMOCCLSDSGVLLSINIKVL 387
DB 301 CEVWGPSPKMLSLKENKEAKVSKREKPVWVINEPAGMOCCLSDSGVLLSINIKVL 360
QY 388 PTWSTPVEP 396
DB 361 PTWSTPVEP 369

RESULT 9
CD4_CERTO STANDARD; PRT; 397 AA.
AC 008336;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3)
DE (Fragment).
GN CD4.
OS Cercopithecus torquatus ayes (Red-crowned mangabey) (Sooty mangabey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OC NCBI_TaxID=9531;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RA MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus." (1992).
RL Eur. J. Immunol. 22:2973-2981(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X73328; CAAS1754.1; -.
CC EMBL: X73327; CAAS1753.1; -.
DR HSP: P01730; IWIQ.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:receptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR007110; Ig_Like.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig_2.
DR SMART: SMART; CD4TCANTIGEN.
DR SMART: SM00406; IGV; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Lipoprotein; Palmitate.
FT DOMAIN 1 369
FT TRANSMEM 370 391
FT DOMAIN <1 98
FT DOMAIN 99 176
FT DOMAIN 177 290
FT DOMAIN 291 347
FT DOMAIN 269 269
FT CARBOHYD 298 298
FT CARBOHYD 14 82
FT DISULFID 128 157
FT DISULFID 301 343
FT LIPID 392 392
FT LIPID 395 395
FT LIPID 397 397
FT NON_TER 397 397
SQ SEQUENCE 397 AA; 44081 MW; 67887397A6B7EA4F CRC64;

```



```

FT DOMAIN <1 98 IG-LIKE V-TYPE.
FT DOMAIN 99 176 IG-LIKE C2-TYPE 1.
FT DOMAIN 177 290 IG-LIKE C2-TYPE 2.
FT DOMAIN 201 347 IG-LIKE C2-TYPE 3.
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 14 82 BY SIMILARITY.
FT DISULFID 128 157 BY SIMILARITY.
FT DISULFID 301 343 BY SIMILARITY.
FT LIPID 392 392 S-palmitoyl cysteine (By similarity).
FT LIPID 395 395 S-palmitoyl cysteine (By similarity).
FT LIPID 20 20 MISSING.
FT VARIANT 43 43 T -> I.
FT VARIANT 86 86 N -> D.
FT VARIANT 96 96 F -> L.
FT VARIANT 173 173 V -> M.
FT VARIANT 316 316 R -> K.
SQ SEQUENCE 397 AA; 43926 MW; 8660B636D2B38A7 CRC64;

Query Match 50.3%; Score 1716; DB 1; Length 397;
Best Local Similarity 89.4%; Pred. No. 6,3e-103;
Matches 330; Conservative 19; Mismatches 20; Indels 0; Gaps 0;

QY 28 VVLGGKGDVVELTCTASQKSIQFPMKNSNOIKILGNOSFLTKGPKSLNDRADSRSLW 87
DB 1 VVLGGKGDVVELACNASQKSKTQFWMKNSKQIKILGNOSFLTKGSSKLSDRADSRSLW 60
QY 88 DQGNPLIILKILKIDSDTYICEVDQKEVQLVFGITANSDDLLOGSLTLTLESP 147
DB 61 DQGSMSIILKILKIDSDTYICEVDQKEVQLVFGITANSDDLLOGSLTLTLESP 120
QY 148 GSSPSVQCRSPRGKNIQSGKTLVSQLELQDSGTCTVLNOKKVEKIDIVLAFQKA 207
DB 121 GSSPSVQCRSPRGKNIQSGKTLVSQLELQDSGTCTVLNOKKVEKIDIVLAFQKA 180
QY 208 SSIIVKKEGEVSEFPLAFVTEKLTGSGELMWQERASSSKSWITFDLKNKEVSVKRV 267
DB 181 SSIIVKKEGEVSEFPLAFVTEKLTGSGELMWQERASSSKSWITFDLKNKEVSVKRV 240
QY 268 QDPKIQMGKPLPLHITLPOALPOVAGSNTLTALEAKTGKILHSEVNLVVMRATOLQKNLT 327
DB 241 QDPKIQMGKPLPLHITLPOALPOVAGSNTLTALEAKTGKILHSEVNLVVMRATOLQKNLT 300
QY 328 CEVWGTPSPKMLSLKLENKAKYKREKPVVNLPEKGMOCILSDSGVLLSNTIKVL 387
DB 301 CEVWGTPSPKMLSLKLENKAKYKREKPVVNLPEKGMOCILSDSGVLLSNTIKVL 360
QY 388 PTWSTPEP 396
DB 361 PTWSTPEP 369

RESULT 10
CD4_SAISC STANDARD; PRT; 457 AA.
AC 029037:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Saimiri.
OC NCBI_TaxId=9521;
RN [1]
RP SEQUENCE FROM N.A.
RA Tatemura M., Hashimoto O.;
RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
-i- FUNCTION: Accessory protein for MHC class-II antigen/T-cell

```

```

CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D86588; BAA1313.1; -.
CC HSSP; P01730; IMBR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR009713; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TcANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
KW SIGNAL
FT CHAIN 1 25
FT 26 457
FT 26 395
FT DOMAIN 396 417
FT TRANSSEM 418 457
FT DOMAIN 26 125
FT DOMAIN 126 202
FT DOMAIN 203 316
FT DOMAIN 317 373
FT CARBOHYD 254 254
FT CARBOHYD 295 295
FT CARBOHYD 324 324
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 327 369
FT LIPID 418 418
FT LIPID 421 421
SQ SEQUENCE 457 AA; 50871 MW; 57BED6344005A015 CRC64;

Query Match 46.3%; Score 1581.5; DB 1; Length 457;
Best Local Similarity 78.0%; Pred. No. 3,2e-94;
Matches 309; Conservative 32; Mismatches 54; Indels 1; Gaps 1;

QY 1 MNRGVPFRHLVLYQALPAATQGNKVYLGKGDVTELTCTASQKSIQFPMKNSNOIK 60
DB 1 MNRGIPFRHLVLYQALPAATVHGKTVVLGKGEVVELCEYSLKKNVPPHMTSQIK 60
QY 11LGNOSFLTKGPKSLNDRADSRSLWDQGNFPLIILKILKIDSDTYICEVDQKEVQL 120
DB 61 ILGQNVFVIRGSKLTDRIIDSKSKSMDRSGFPLIDAIIDSESYTICEVSEKEVEL 120
QY 121 LVFGLTANSDDLLOGSLTLTLESPGSSPVQCRSPRGKNIQSGKTLVSQLELQDSG 180
DB 121 QVFGTLTANSDDLLOGSLTLTLESPGSSPVSECTSPRKRIKRGKTLVSQLELQDSG 180
QY 181 TWTCVTVQNGKXVEFKIDIVLAFQKASSIVYKKEGEVSEFPLAFVTEKLTGSGELMW 240
DB 181 TWTCVTVQHLLEV-FEINIVLAFQQAASSIVYKKEGEVSEFPLAATTLTGSGELCW 239
QY 241 QABRASSSKSWITFDLKNKEVSVKRVTPDKLQMGKPLPLHITLPOALPOVAGSNTLA 300

```



```

Db      240 QABRASSKSWITFLNLTKEVYKVLVTQPKLRMGKGLPLHLTLAQLPQYAGSGNFTLA 299
Qy      301 LEAKGKGLHGVNLYMRAVTOLOKNTCEVWGPTSKMLSLKLENKEAKVSKREKPVAV 360
Db      300 LKGRKGLHGVNLYMRAVTOLOKNTCEVWGPTSKMLSLKLENKEAKVSKREKPVAV 359
Qy      361 LNPEAGMOCCLSDSGVLLSENIKVLPTWSTPEVP 396
Db      360 LNPEAGMOCCLSDSGVLLSENIKVLPTWSTPEVP 395

RESULT 11
GCL_HUMAN STANDARD; PRT; 330 AA.
AC POL557;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE IG gamma-1 chain C region.
GN IGHG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82274238; PubMed=6287432;
RA Ellson J.W., Berson B.J., Hood L.E.;
RT "The nucleotide sequence of a human immunoglobulin C gamma1 gene.";
RL Nucleic Acids Res. 10:4071-4079(1982).
RN [2]
RP SEQUENCE OF 1-135 (MYELOMA PROTEIN EU).
RX MEDLINE=71064024; PubMed=5489771;
RA Cunningham B.A., Rutishauser U., Gall W.E., Gottlieb P.D.,
RA Waxdal M.J., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. VII. Amino
RT acid sequence of heavy-chain cyanogen bromide fragments H1-H4.";
RL Biochemistry 9:3161-3170(1970).
RN [3]
RP SEQUENCE OF 136-329 (EU).
RX MEDLINE=71064025; PubMed=5530842;
RA Rutishauser U., Cunningham B.A., Bennett C., Konigsberg W.H.,
RA Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. 8. Amino
RT acid sequence of heavy-chain cyanogen bromide fragments H5-H7.";
RL Biochemistry 9:3171-3181(1970).
RN [4]
RP SEQUENCE (MYELOMA PROTEIN NIE).
RX MEDLINE=77070269; PubMed=826475;
RA Poncetngl H., Hilschmann N.;
RT "The rule of antibody structure. The primary structure of a
RT monoclonal IgG1 immunoglobulin (myeloma protein Nie). III. The
RT chymotryptic peptides of the H-chain, alignment of the tryptic
RT peptides and a discussion of the complete structure.";
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1571-1604(1976).
RN [5]
RP SEQUENCE (MYELOMA PROTEIN KOL), AND DISULFIDE BONDS.
RX MEDLINE=83289131; PubMed=6884994;
RA Schmidt W.E., Jung H.-D., Palm M., Hilschmann N.;
RT "Three-dimensional structure determination of antibodies. Primary
RT structure of crystallized monoclonal immunoglobulin IgG1 KOL, I.";
RL Hoppe-Seyler's Z. Physiol. Chem. 364:713-747(1983).
RN [6]
RP DISULFIDE BONDS.
RX MEDLINE=71064027; PubMed=4923144;
RA Gall W.E., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. X.
RT interchain disulfide bonds.";
RL Biochemistry 9:3188-3196(1970).
RN [7]
RP DISULFIDE BONDS.
RX MEDLINE=77070267; PubMed=1002129;

```

```

RA Dreker L., Schwarz J., Reichel W., Hilschmann N.;
RT "Rule of antibody structure. The primary structure of a monoclonal
RT IgG1 immunoglobulin (myeloma protein Nie). I: Purification and
RT characterization of the protein, the L- and H-chains, the
RT cyanogen bromide cleavage products, and the disulfide bridges.";
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540(1976).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RX MEDLINE=81208100; PubMed=7236608;
RA Deisenhofer J.;
RT "Crystallographic refinement and atomic models of a human Fc fragment
RT and its complex with fragment B of protein A from Staphylococcus
RT aureus at 2.9- and 2.8-A resolution.";
RL Biochemistry 20:2361-2370(1981).
CC -I- MISCELLANEOUS: Nie has the G1M(17) allotypic marker, 97-K, and the
CC G1M(1) markers, 239-D and 241-L. KOL and EU sequences have the
CC G1M(3) marker and the G1M (non-1) markers.
CC -I- MISCELLANEOUS: Nie also differs in the amidation states of
CC 35, 116, 198, 269 and 272.
CC -I- MISCELLANEOUS: EU also differs in the amidation states of residues
CC 155, 166, 177, 195, 198, 269, and 272 and in the order of residues
CC 268-272.
CC -I- MISCELLANEOUS: KOL also differs in the amidation states of
CC residues 198, 267 and 272.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: J00228; AAC82527.1; ALT_INT.
DR PIR: A93433; GHU.
DR PDB: 1FC1; 15-JUL-92.
DR PDB: 1FC2; 15-JUL-92.
DR PDB: 1A77; 12-NOV-97.
DR PDB: 1A5B; 09-FEB-00.
DR PDB: 1D5T; 09-FEB-00.
DR PDB: 1D6V; 04-OCT-00.
DR PDB: 1DN2; 17-MAY-00.
DR PDB: 1E4K; 06-JUN-01.
DR PDB: 1FCC; 20-JUL-95.
DR PDB: 1H2H; 12-JUN-02.
DR PDB: 1I7Z; 08-AUG-01.
DR PDB: 1I1S; 16-MAY-01.
DR PDB: 1I1X; 16-MAY-01.
DR PDB: 1L6X; 10-APR-02.
DR PDB: 2RCS; 12-NOV-97.
DR GeneW; HGNC:5525; IGHG1.
DR MIM; 147100; -.
DR GO; GO:0005624; C:membrane fraction; NAS.
DR GO; GO:0003823; F:antigen binding; TAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig-cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW 3D-structure.
KW NON_TER
FT 1 98 CH1.
FT DOMAIN 99 110 HINGE.
FT DOMAIN 111 223 CH2.
FT DOMAIN 224 330 CH3.
FT DISULFID 27 83
FT DISULFID 103 103 INTERCHAIN (WITH LIGHT CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH HEAVY CHAIN).

```



```

FT DISULFID 144 204
FT DISULFID 250 308
FT CARBOHD 180 180
FT MOD RES 330 330
FT VARIANT 97 97
FT VARIANT 239 239
FT VARIANT 241 241
FT STRAND 122 126
FT HELIX 130 134
FT TURN 136 137
FT STRAND 141 147
FT STRAND 157 162
FT TURN 163 164
FT STRAND 165 166
FT TURN 168 171
FT STRAND 176 179
FT TURN 180 181
FT STRAND 182 190
FT HELIX 193 197
FT TURN 198 199
FT STRAND 202 207
FT TURN 209 210
FT STRAND 215 219
FT STRAND 227 227
FT STRAND 230 234
FT HELIX 238 242
FT STRAND 245 256
FT STRAND 260 265
FT STRAND 270 270
FT STRAND 274 276
FT STRAND 280 281
FT TURN 283 284
FT STRAND 287 296
FT HELIX 297 301
FT TURN 302 303
FT STRAND 305 312
FT TURN 313 314
FT TURN 316 317
FT STRAND 320 325
SQ SEQUENCE 330 AA; 36106 MW; 3770EB106CC2FA3D CRC64;

Query Match 37.1%; Score 1266; DB 1; Length 330;
Best Local Similarity 96.7%; Pred. No. 3.5e-74;
Matches 236; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

OY 382 SNIKVLPWSTPVEPKSCDKTHHTCPCPAPPELLGGPSVFLFPPPKKDTLMISRTPEVTCV 441
DB 90 SNTKV---DKKVEPKSCDKTHHTCPCPAPPELLGGPSVFLFPPPKKDTLMISRTPEVTCV 145
OY 442 VVDVSHEDPEVKFMNVYDGVENNAKTRREOYNSTTRVSVLTVLHODMLNKEYKCK 501
DB 146 VVDVSHEDPEVKFMNVYDGVENNAKTRREOYNSTTRVSVLTVLHODMLNKEYKCK 205
OY 502 VSNKALPAPIEKTKISKAGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVME 561
DB 206 VSNKALPAPIEKTKISKAGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVME 265
OY 562 SNGQPENNYKTTTPYLDSDGSFFLYSKLTVDKSRWQQGNVSCSVMHEALHNHYTQKSLS 621
DB 266 SNGQPENNYKTTTPYLDSDGSFFLYSKLTVDKSRWQQGNVSCSVMHEALHNHYTQKSLS 325
OY 622 LSPG 625
DB 326 LSPG 329

```

```

DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE IG gamma-3 chain C region (Heavy chain disease protein) (HDC).
GN IGHG3.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE (DISEASE PROTEIN WIS).
RX MEDLINE=81021546; PubMed=6774747;
RA Frangione B., Rosenwasser E., Prelli F., Franklin E.C.;
RT "Primary structure of human gamma 3 immunoglobulin deletion mutant:
RN gamma 3 heavy-chain disease protein WIS."
RL Biochemistry 19:4304-4308(1980).
RP REVISIONS TO 12-97 (PROTEIN WIS).
RX MEDLINE=77118561; PubMed=402363;
RA Michaelson T.E., Frangione B., Franklin E.C.;
RT "Primary structure of the 'hinge' region of human IgG3. Probable
RL quadruplication of a 15-amino acid residue basic unit.",
RN J. Biol. Chem. 252:883-889(1977).
RP REVISIONS TO 59-289 (PROTEIN WIS/DISEASE PROTEIN ZUC).
RX MEDLINE=77021516; PubMed=823945;
RA Wolfenstein-Todel C., Frangione B., Prelli F., Franklin E.C.;
RT "The amino acid sequence of 'heavy chain disease' protein ZUC.
RL Structure of the Fc fragment of immunoglobulin G3."
RN Biochem. Biophys. Res. Commun. 71:907-914(1976).
RP SEQUENCE FROM N.A. (DISEASE PROTEIN OWM).
RX MEDLINE=82247835; PubMed=6808505;
RA Alexander A., Steinmetz M., Barilaute D., Frangione B.,
RN Franklin E.C., Hood L., Buxbaum J.N.;
RT "Gamma heavy chain disease in man: cDNA sequence supports partial
RL gene deletion model."
RN Proc. Natl. Acad. Sci. U.S.A. 79:3260-3264(1982).
CC -1- SUBUNIT: Dimer linked by 12 disulfide bonds; it has an extra
interchain disulfide bond at position 7 in addition to the 11
normally present in the hinge region.
CC -1- MISCELLANEOUS: The heavy chain disease protein WIS is shown.
CC -1- MISCELLANEOUS: The sequence of residues 42-76 was taken from the
Ref.2.
CC -1- MISCELLANEOUS: Disease protein WIS is lacking most of the V region
and all of the CH1 region.
CC -1- MISCELLANEOUS: Disease protein ZUC lack most of the V region, all
of the CH1 region, and part of the hinge compared with normal
gamma-3 heavy chains.
CC -1- MISCELLANEOUS: Disease protein OWM may represent an allelic form
or another gamma chain subclass.
CC -1- MISCELLANEOUS: The hinge region in gamma-3 chains is about four
times as long as in other gamma chains and contains three
identical 15-residue segments preceded by a similar 17-residue
segment (12-28).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL, J00231; AA52805.1; ALT_SEQ.
DR HSSP; P01857; IFC1.
DR Genew; HGNC:5527; IGHG3.
DR MIM; 147120; -.
DR GO; GO:0005624; C:membrane fraction; NMS.
DR GO; GO:0003823; P:antigen binding; TAS.
DR GO; GO:0006955; P:immune response; NMS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.

```



```

DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig1; 2.
DR SMART: SM00407; Igcl; 1.
DR PROSITE: PS50835; IG_LIKE; 2.
DR PROSITE: PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat;
KW Pyridoxal carboxylic acid.
FT DOMAIN 12 73 HINGE.
FT DOMAIN 74 183 CH2.
FT DOMAIN 184 289 CH3.
FT REPEAT 229 43
FT REPEAT 44 58
FT REPEAT 59 73
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT CARBOHYD 6 6 N-LINKED (GLCNAC...).
FT DISULFID 7 7 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 27 7 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 27 27 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 33 33 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 39 39 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 42 42 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 48 48 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 54 54 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 57 57 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 63 63 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 69 69 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT DISULFID 72 72 INTERCHAIN (WITH HEAVY CHAIN DIMER).
FT CARBOHYD 140 140 N-LINKED (GLCNAC...).
FT MOD_RES 290 290 REMOVED POST-TRANSLATIONALLY.
FT VARIANT 126 127 QV -> EB (IN ZUC).
FT VARIANT 134 134 /FTID=VAR 003890.
FT VARIANT 139 139 P -> L (IN OMM).
FT VARIANT 182 182 F -> Y (IN OMM).
FT VARIANT 227 227 /FTID=VAR 003892.
FT VARIANT 227 227 T -> A (IN OMM).
FT VARIANT 227 227 /FTID=VAR 003893.
FT VARIANT 227 227 S -> N (IN OMM).
FT VARIANT 227 227 /FTID=VAR 003894.
FT VARIANT 227 227 MISSING (IN ZUC).
FT VARIANT 279 279 /FTID=VAR 003895.
FT VARIANT 279 279 F -> Y (IN OMM).
FT SEQUENCE 290 AA; 32331 MW; E69CB035705B2F46 CRC64;
SQ
Query Match 33.7%; Score 1151; DB 1; Length 290;
Best Local Similarity 90.5%; Pred. No. 6; 9e-67;
Matches 209; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

```

```

DE Ig gamma-2 chain C region.
GN IGH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE OF 2-326 FROM N.A.
RX MEDLINE=82197621; PubMed=6804948;
RA Eljison J.W., Hood L.E.;
RT "Linkage and sequence homology of two human immunoglobulin gamma
RT heavy chain constant region genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1984-1988(1982).
RN [2]
RP SEQUENCE OF 88-115 FROM N.A.
RC TISSUE=Petal liver;
RX MEDLINE=83001943; PubMed=681139;
RA Takahashi N., Ueda S., Obata M., Nikaide T., Nakai S., Honjo T.;
RT "Structure of human immunoglobulin gamma genes: implications for
RT evolution of a gene family.";
RL Cell 29:671-679(1982).
RN [3]
RP SEQUENCE OF 99-177 AND 310-326 FROM N.A.
RC TISSUE=Petal liver;
RX MEDLINE=84235992; PubMed=6329676;
RA Krawinkel U., Rabbits T.H.;
RT "Comparison of the hinge-coding segments in human immunoglobulin gamma
RT heavy chain genes and the linkage of the gamma 2 and gamma 4 subclases
RT genes.";
RL EMBO J. 1:403-407(1982).
RN [4]
RP SEQUENCE OF 1-325 (MYELOMA PROTEIN TIL).
RX MEDLINE=81007873; PubMed=6774012;
RA Wang A.-C., Tung E., Fudenberg H.H.;
RT "The primary structure of a human IgG2 heavy chain: genetic,
RT evolutionary, and functional implications.";
RL J. Immunol. 125:1048-1054(1980).
RN [5]
RP SEQUENCE OF 1-85 AND 132-325 (MYELOMA PROTEIN ZIE).
RX MEDLINE=80001357; PubMed=113060;
RA Connell G.E., Parr D.M., Hofmann T.;
RT "The amino acid sequences of the three heavy chain constant region
RT domains of a human IgG2 myeloma protein.";
RL Can. J. Biochem. 57:758-767(1979).
RN [6]
RP SEQUENCE OF 238-275 (ZIE).
RX MEDLINE=80114419; PubMed=118920;
RA Hofmann T., Parr D.M.;
RT "A note of the amino acid sequence of residues 381-391 of human
RT immunoglobulin gamma chains";
RL Mol. Immunol. 16:923-925(1979).
RN [7]
RP REVISIONS TO 25; 59; 60 AND 264-268 (ZIE).
RA Hofmann T., Parr D.M.;
RT submitted (MAR-1980) to the PIR data bank.
RN [8]
RP SEQUENCE OF 1-121 (DOT).
RX MEDLINE=95255298; PubMed=7737190;
RA Scoppini M., Bellotti V., Negri A., Merlini G., Garver F., Ferri G.;
RT "Characterization of the two unique human anti-flavin monoclonal
RT immunoglobulins";
RL Eur. J. Biochem. 228:886-893(1995).
RN [9]
RP DISULFIDE BONDS.
RX MEDLINE=72033500; PubMed=4940472;
RA Milestein C., Frangione B.;
RT "Disulphide bridges of the heavy chain of human immunoglobulin G2.";
RL Biochem. J. 121:217-225(1971).
RN [10]
RP DISULFIDE BONDS.
RX MEDLINE=69064124; PubMed=5782707;
RA Frangione B., Milestein C., Pink J.R.L.;
RT "Structural studies of immunoglobulin G.";

```



RL Nature 221:145-148 (1969).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: J00230; AAB59393.1; -.  
DR PIR: A83906; G2HU.  
DR HSSP: P01857; 1FC1.  
DR Genew: HGNC:5526; IGHG2.  
DR MIM: 147110; -.  
DR GO: GO:0005624; C:membrane fraction; NAS.  
DR GO: GO:0003823; F:antigen binding; NAS.  
DR GO: GO:0006955; P:immune response; NAS.  
DR InterPro: IPR007110; IG-like.  
DR InterPro: IPR003597; IG\_c1.  
DR InterPro: IPR003006; IG\_MHC.  
DR Pfam: PF00047; ig; 3.  
DR SMART: SM00407; IGc1; 2.  
DR PROSITE: PS00835; IG\_LIKE; 3.  
DR PROSITE: PS00290; IG\_MHC; 2.  
KM Immunoglobulin domain; Immunoglobulin C region.  
FT NON TER 1 1  
FT DOMAIN 1 98 CH1.  
FT DOMAIN 99 110 HINGE.  
FT DOMAIN 111 219 HINGE.  
FT DOMAIN 220 326 CH3.  
FT DISULFID 14 14 INTERCHAIN (WITH A LIGHT CHAIN).  
FT DISULFID 27 83 INTERCHAIN (WITH A HEAVY CHAIN).  
FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).  
FT DISULFID 103 103 INTERCHAIN (WITH A HEAVY CHAIN).  
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).  
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).  
FT DISULFID 140 200  
FT DISULFID 246 200  
FT SITE 156 156 AT OR NEAR THE COMPLEMENT-BINDING SITE.  
FT MOD RES 326 326 REMOVED POST-TRANSLATIONALLY (PROBABLY).  
FT VARIANT 60 60 S-> A (IN MYELOMA PROTEIN TIL & ZIE).  
FT CONFLICT 109 109 /FTID=VAR\_003889.  
SQ SEQUENCE 326 AA; 35884 MW; 83108786869CFC64; CRC64;  
-----  
Query Match 33.7%; Score 1150; DB 1; Length 326;  
Best Local Similarity 88.9%; Pred. No. 9,3e-67;  
Matches 217; Conservative 9; Mismatches 10; Indels 8; Gaps 3;  
-----  
QY 382 SNIKLPTWSTPVEKSCDKTHTCPCPAPELLGSPVFLFPKPKDTLMTSRPEVTCV 441  
DB 90 SNTKVDK-----VERKCV-----CPCPCAP- VAGPSVFLFPKPKDTLMTSRPEVTCV 141  
-----  
QY 442 VVDVSHEDPEVKNYVDGVEVHNATKYPREBOYNSTYRVSVLTVLHQMVLNGEYKCK 501  
DB 142 VVDVSHEDPEVQFNMYVDGVEVHNATKYPREBOYNSTYRVSVLTVLHQMVLNGEYKCK 201  
-----  
QY 502 VSNKALPPIETKTSKAGQPREPOVYTLPSRDLTKNOVSLTCLVNGFPPSDIAVWE 561  
DB 202 VSNKGLPPIETKTSKAGQPREPOVYTLPSRDLTKNOVSLTCLVNGFPPSDIAVWE 261  
-----  
QY 562 SNGQENNYKTTPTVLDSDGSFFLYSKLTVDSKRMQGNVSCSVMEALNNHTQKSL 621  
DB 262 SNGQENNYKTTPTVLDSDGSFFLYSKLTVDSKRMQGNVSCSVMEALNNHTQKSL 321  
-----  
QY 622 LSPG 625  
DB 322 LSPG 325

ID CD4\_RABIT STANDARD; PRT; 459 AA.  
AC P46630;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen  
DE T4/Leu-3).  
GN CD4.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OC NCBI\_TaxID=9986;  
RX MEDLINE=92390370; PubMed=1518821;  
RA Hague B.F., Sawasdi Kosol S., Brown T.J., Lee K., Recker D.P.,  
RA Kindt T.O.;  
RT "CD4 and its role in infection of rabbit cell lines by human  
RT immunodeficiency virus type 1."  
RL Proc. Natl. Acad. Sci. U.S.A. 89:7963-7967(1992).  
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell  
CC receptor interaction. May regulate T-cell activation.  
CC -1- SUBUNIT: Associates with p56-lck (by similarity).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M92840; AAA31198.1; -.  
DR PIR: A46254; A46254.  
DR HSSP: P01730; 1WBR.  
DR GO: GO:0042101; C:T-cell receptor complex; ISS.  
DR GO: GO:0015026; F:coreceptor activity; ISS.  
DR GO: GO:0042289; F:MHC class II protein binding; ISS.  
DR GO: GO:0006955; P:immune response; ISS.  
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.  
DR GO: GO:0030217; P:T-cell differentiation; ISS.  
DR GO: GO:0045058; P:T-cell selection; ISS.  
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.  
DR InterPro: IPR000973; CD4\_TcAg.  
DR InterPro: IPR007110; IG-like.  
DR InterPro: IPR003596; IG\_v.  
DR Pfam: PF00047; ig; 2.  
DR PRINTS: PR00692; CD4TCANTIGEN.  
DR SMART: SM00406; IGv; 1.  
DR PROSITE: PS00835; IG\_LIKE; 1.  
KM Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;  
KM Immune response; Repeat; Signal; Lipoprotein; Palmitate.  
FT SIGNAL 1 25  
FT CHAIN 26 459  
FT DOMAIN 26 396  
FT TRANSLEM 397 419  
FT DOMAIN 420 459  
FT DOMAIN 26 129  
FT DOMAIN 130 308  
FT DOMAIN 209 318  
FT DOMAIN 319 374  
FT CARBOHYD 299 299  
FT DISULFID 41 113  
FT DISULFID 329 370  
FT LIPID 420 420  
FT LIPID 423 423  
SQ SEQUENCE 459 AA; 50886 MW; B323311CDB40013D CRC64;  
-----  
Query Match 33.7%; Score 1149; DB 1; Length 459;  
Best Local Similarity 56.9%; Pred. No. 1.7e-66;  
-----







```

RP SEQUENCE FROM N.A.
RC STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=94378217; PubMed=8091416;
RA Gorman S.D., Frewin M.R., Cobbold S.P., Waldmann H.;
RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
  alpha antigens."
RL Tissue Antigens 43:184-188(1994).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in macrophages and a subset of
CC T lymphocytes.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L06130; AB02295.1; -
DR EMBL; X68565; -; NOT_ANNOTATED_CDS.
DR HSSP; P01730; 1MBR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig-TcAg.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 3.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 24
FT CHAIN 25 463
FT DOMAIN 25 401 T-CELL SURFACE GLYCOPROTEIN CD4.
FT TRANSMEM 402 423
FT DOMAIN 424 463 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 125 124 IG-LIKE V-TYPE.
FT DOMAIN 125 211 IG-LIKE C2-TYPE 1.
FT DOMAIN 212 321 IG-LIKE C2-TYPE 2.
FT DOMAIN 322 378 IG-LIKE C2-TYPE 3.
FT DISULFID 41 109
FT DISULFID 332 374 BY SIMILARITY.
FT DISULFID 424 424 S-palmitoyl cysteine (By similarity).
FT LIPID 427 427 S-palmitoyl cysteine (By similarity).
FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 389 389 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 463 AA; 51639 MW; 95805170CB44A833 CRC64;

```

```

Query Match 33.4%; Score 1141; DB 1; Length 463;
Best Local Similarity 57.6%; Pred. No. 5,6e-66;
Matches 236; Conservative 63; Mismatches 93; Indels 18; Gaps 6;

```

```

QY 1 MNRGVPFPHLLVLTALLPATQGNKVLKGGDTVLTCTASGSKSIQTHWKNNSNIK 60
DB 1 MNRGAPFPHLLVLTALLPATQGNKVLKGGDTVLTCTASGSKSIQTHWKNNSNIK 60

```

```

QY 61 ILNGQSFLLTKGPKLNDRADSRRLWDQGNFLLINKLIEDSDTYICEVEDQKEEVQL 120
DB 61 ILNGQSFLLTKGPKLNDRADSRRLWDQGNFLLINKLIEDSDTYICEVEDQKEEVQL 119
QY 121 LVFGLTL-----NSDTHLLOGSLTITLSEPPSSPSVQCRSRGNIOGKTLNV 171
DB 120 LVFNLTKMKDSSGSSGNRLRLOGQQLTTLLEPSSSPSVQKGGNNSKHGQNLSTL 179
QY 172 SQELDQSGTWCTCVLONOKKVEFKIDIVLAFQKASSIYKKGEQVFPFPAFTVEK 231
DB 180 SWEPLDQSGTWCTCVLONOKKVEFKIDIVLAFQKASSIYKKGEQVFPFPAFTVEK 239
QY 232 LYSGLSLMWAERASSSSKSWITFDLKNKEVSVKRVTPDKLQWKKPLHLTLQALPQY 291
DB 240 LV-GEIRWQAQAGASSSLWISFTLENRKLSMKEAHAFLKIQMKESLPLRFTLPQVLSRY 297
QY 292 AGSNLTLLAEKATGKHQEVNLYVMKATLOKULTEWGPSPKMLSLKENKAVK 351
DB 298 AGSGILTLNL-AK-GTLQEVNLYVMKANSNNLTCEVIGPSPSPELTLSINLEQAKV 355
QY 352 SKREKPYVWNPBAGWMOCLSDSGVLTLSNIVKLPWSPVPEPKSCDK 401
DB 356 SKQKLVWVVDPEBGTWQCLSDKDKVLLASSLVN---SSPVVYKSWPK 401

```

```

RESULT 17
CD4_RAT ID CD4_RAT STANDARD; PRT; 457 AA.
AC P05540;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
  T4/Leu-3) (W3/25 antigen).
GN CD4.
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=8717535; PubMed=3104900;
RA Clark S.D., Jeffries W.A., Barclay A.N., Gagnon J., Williams A.F.;
RT "Peptide and nucleotide sequences of rat CD4 (W3/25) antigen:
  evidence for derivation from a structure with four
  immunoglobulin-related domains."
RL Proc. Natl. Acad. Sci. U.S.A. 84:1649-1653(1987).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 210-393.
RX MEDLINE=93262437; PubMed=8493535;
RA Brady R.L., Dodson E.J., Dodson G.G., Lange G., Davis S.J.,
  Williams A.F., Barclay A.N.;
RT "Crystal structure of domains 3 and 4 of rat CD4: relation to the
  NH2-terminal domains."
RL Science 260:979-983(1993).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like V-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15768; AAA40901.1; -
DR PIR; A27449; A27449.
DR PIR; ICID; 15-JUL-93.

```



```
DR GlycoSiteDB; P05540; C: T-cell receptor complex; ISS.
DR GO; GO:0042101; C: T-cell receptor activity; ISS.
DR GO; GO:0015026; F: coreceptor activity; ISS.
DR GO; GO:0042289; F: MHC class II protein binding; ISS.
DR GO; GO:0006955; P: immune response; ISS.
DR GO; GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P: T-cell differentiation; ISS.
DR GO; GO:0045058; P: T-cell selection; ISS.
DR GO; GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcRg.
DR InterPro; IPR007110; Ig-ILike.
DR InterPro; IPR003599; Ig.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00409; Ig; 2.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure.
FT CHAIN 1 27 T-CELL SURFACE GLYCOPROTEIN CD4.
FT SIGNAL 28 457 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 28 394 POTENTIAL.
FT TRANSMEM 395 417 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 418 457 IG-LIKE V-TYPE.
FT DOMAIN 128 127 IG-LIKE C2-TYPE 1.
FT DOMAIN 207 206 IG-LIKE C2-TYPE 2.
FT DOMAIN 317 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 43 111 BY SIMILARITY.
FT DISULFID 158 187 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 418 418 S-palmitoyl cysteine (By similarity).
FT LIPID 421 421 S-palmitoyl cysteine (By similarity).
FT STRAND 213 217
FT TURN 218 219
FT STRAND 222 225
FT STRAND 235 243
FT STRAND 252 258
FT TURN 259 260
FT STRAND 261 265
FT STRAND 274 275
FT STRAND 278 278
FT TURN 279 280
FT STRAND 282 285
FT HELIX 290 292
FT STRAND 294 301
FT STRAND 306 319
FT STRAND 325 331
FT STRAND 338 344
FT TURN 345 346
FT STRAND 349 353
FT STRAND 357 361
FT STRAND 367 374
FT TURN 375 376
FT STRAND 377 385
SQ SEQUENCE 457 AA; 51437 MW; 477BE157D30954C1 CRC64;

Query Match 29.3%; Score 999; DB 1; Length 457;
Best Local Similarity 52.9%; Pred. No. 6,9e-57;
Matches 207; Conservative 61; Mismatches 115; Indels 8; Gaps 5;
```

```
Db 121 ELMEFRVTENPGTRLLQGQSLLTLIDSNPKVSDPPEICCKHSSNIVKDSKAFSTSLRIQ 180
Qy 178 DSGTTCYTLQOKKVEFKIDIVLAFQKASSIVYKKEGEVFEFPPLAFYKLTSGGE 237
Db 181 DSGINCVTLTNQKHSFPMKLSVIGFASSTSTAYAKSREBSAEFSPPLNGESL--QGE 238
Qy 238 LMQAERASSSKSWITFDLKNKEVSKRVTOODPKLQMGKKLPLHLTLPOALPOVAGSGNL 297
Db 239 LMKKEKAPSSQSWITFSLKNOKSVOKSTSNPKQLSETLPLTLQIPOVSLQFAGSGNL 298
Qy 298 TLAEAKTGKLRQEVNLVYMRATQLOKN-LTCEVWGPTSPKXLSLKLKENKAKYKREK 356
Db 299 TLTLD--RGILYQEVNLVYMKVTOQPSNTLTCEVWGPTSPKXRLILKQENQEARVSRQEK 356
Qy 357 PYVYVNPENPGMVOCLISDSGQVLESNIKVL 387
Db 357 VIQVAPENPAGVQCLISREGEVMDSKIQVL 387

RESULT 18
CD4_MOUSE STANDARD; PRT; 457 AA.
ID CD4_MOUSE
AC P06332;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3) (T-cell differentiation antigen L3T4).
GN CD4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87018845; PubMed=3094146;
RA Tourville B., Gorman S.D., Field E.H., Hunkapiller T., Parnes J.R.;
RT "Isolation and sequence of L3T4 complementary DNA clones: expression
RT in T cells and brain.";
RL Science 234:610-614 (1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87115821; PubMed=3027575;
RA Littman D.R., Gettner S.N.;
RT "Unusual Intron in the immunoglobulin domain of the newly isolated
RT murine CD4 (L3T4) gene.";
RL Nature 325:453-455 (1987).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=88152875; PubMed=3326818;
RA Parnes J.R., Hunkapiller T.;
RT "L3T4 and the immunoglobulin gene superfamily: new relationships
RT between the immune system and the nervous system.";
RL Immunol. Rev. 100:109-127 (1987).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=88041159; PubMed=2823269;
RA Gorman S.D., Tourville B., Parnes J.R.;
RT "Structure of the mouse gene encoding CD4 and an unusual transcript
RT in brain.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648 (1987).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=98112780; PubMed=9445485;
RA Anasetti L.M.A., Oeltjen J.C., Schwartz S., Zhang Z., Muzny D.M.,
RA Lu J., Correll J.H., Chinnault A.C., Belmont J.W., Miller W.,
RA Gibbs R.A.;
RT "Comparative sequence analysis of a gene-rich cluster at human
RT chromosome 12p13 and its syntenic region in mouse chromosome 6.";
RL Genome Res. 8:29-40 (1998).
RN [6]
```



```

RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rabe S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bock S.A., McEwan P.J., McKernan K.J., Malik J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield J.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

[7]
RP SEQUENCE OF 27-43.
RX MEDLINE=8616694; PubMed=3082751;
RA Classon B.J., Tsagaratos J., Kirsbaum L., Maddox J., McKay C.R.,
RA Brandon M., McKenzie I.F.C., Walker I.D.;
RT "The L12A antigen in mouse and the sheep equivalent are
RT immunoglobulin-like".
RL Immunogenetics 23:129-132 (1986).

[8]
RP DISULFIDE BONDS.
RX MEDLINE=86233454; PubMed=3086886;
RA Classon B.J., Tsagaratos J., McKenzie I.F.C., Walker I.D.;
RT "Partial primary structure of the T4 antigens of mouse and sheep:
RT assignment of intrachain disulfide bonds."
RL Proc. Natl. Acad. Sci. U.S.A. 83:4499-4503 (1986).

CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lick (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: Type I membrane protein.
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P06332-1; Sequence=Displayed;
CC Name=2; Synonyms=Brain-specific;
CC IsoId=P06332-2; Sequence=VSP_002489;
CC -1- SIMILARITY: Contains 3 Immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 Immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M36850; AAA39401.1; -
CC EMBL; M13816; AAA37267.1; -
CC EMBL; X04836; CAA28539.1; -
CC EMBL; M36851; AAA39402.1; -
CC EMBL; M17080; AAA37403.1; -
CC EMBL; M17078; AAA37403.1; JOINED.
CC EMBL; M17079; AAA37403.1; JOINED.
CC EMBL; AC002397; AAC36010.1; -
CC EMBL; BC039137; AAC39137.1; -
CC PIR; A02110; RWMST4.
CC DR HSSP; P01730; 1WBR.
CC MGD; MGI:88335; Cd4.
CC GO; GO:0042101; C:T-cell receptor complex; ISS.
CC GO; GO:0015026; F:coreceptor activity; ISS.

```

```

DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcRg.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SMO0406; IGV_1.
DR PROSITE; PS50835; IG_LIKE_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate;
KW Alternative splicing.
FT SIGNAL 1 26
FT CHAIN 27 457 T-CELL SURFACE GLYCOPROTEIN CD4.
FT DOMAIN 27 394 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 395 417 POTENTIAL.
FT DOMAIN 418 457 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 27 128 IG-LIKE V-TYPE.
FT DOMAIN 129 207 IG-LIKE V-TYPE.
FT DOMAIN 208 317 IG-LIKE C2-TYPE 1.
FT DOMAIN 318 374 IG-LIKE C2-TYPE 2.
FT CARBOHYD 187 187 IG-LIKE C2-TYPE 3.
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 42 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 159 188
FT DISULFID 328 370
FT LIPID 418 418 S-palmitoyl cysteine (By similarity).
FT LIPID 421 421 S-palmitoyl cysteine (By similarity).
FT VARSPLIC 1 240 Missing (in isoform 2).
FT /FTId=VSP_002489.
FT /FTId=VSP_002489.
SO SEQUENCE 457 AA; 51296 MM; 1818A7527C800F33 CRC64;

Query Match 29.1%; Score 993; DB 1; Length 457;
Best Local Similarity 53.8%; Pred. No. 1.7e-56;
Matches 211; Conservative 62; Mismatches 109; Indels 10; Gaps 7;

QY 1 MNRGVPRRH-LLVLTQLALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHKNKNQI 59
DB 1 MCRATSLRRLLILLQLSOLLAVTQGLTVLGKGESEALPCSSQKRTVFTWKFSDOR 60
QY 60 KIINGNS-SFLTGS--PSKLNDRADRSRLMDQSNFPLIIKNKIENSDPYICEVEDOKE 116
DB 61 KIISQHGKGVILIRGSPSPQ--DRPDSKKGAMKSPFLIINKLKMEDSQTITCELENRKE 119
QY 117 EVQVLVEGLTRANSPDTHLQGSQSLTLTLDS--PPGSSPEVQCRSPRGKNIQGGKTLVSQLE 175
DB 120 EVELMWVKVTFSPETSLLQGSQSLTLTLDSNSKVSNPFLTECKHKKGKVVSGSKVLSMNL 179
QY 176 LQDSGTCTCTVLQNKVVEKIDIVLAFQKASSIVYKKEGBOVESFPPLAFVYEKLTGS 235
DB 180 VQSDPFNQCVTYLLDQKKNMGMTLSVGFQSTAITAYKSGESAEFPPLFAEE--NGW 237
QY 236 GELMWQNERASSKSWTFPLDKXKEVSVKRVTDOPKLOMKKKLPLHLTLPOALPOVAGSG 295
DB 238 GELMWKAKKDSFPQVWTSFSIKNEVSVQKSTDKLQKLETPPLTLTKIPVSLQFPGSG 297
QY 296 NLTLALFAKTKGLHQEVNLVVMRATQLOKNLTCEVMGPTSPKMLSLKLENKAKVSKRE 355
DB 298 NLTLTLTD--KGTLHQEVNLVVMKVAQNLNLTCEVMGPTSPKMKTLTKQENGEARVSEEQ 355
QY 356 KPVVNLNPEAGMWQCLISDSGQVLLSESNIVYL 387
DB 356 KVVQVNVAPETGLWQCLISDSGQVVMDSRIQVL 387

RESULT 19
GC3M_MOUSE

```







DR PIR, B02159; GIMSM.  
 DR PDB; 15C8; 23-MAR-99.  
 DR PDB; 1AE6; 18-MAR-98.  
 DR PDB; 1CL7; 12-JAN-00.  
 DR PDB; 1FL1; 06-FEB-01.  
 DR PDB; 1P58; 29-DEC-99.  
 DR PDB; 1KC5; 24-JUL-02.  
 DR PDB; 1KCR; 11-MAY-02.  
 DR PDB; 25C8; 09-JUL-99.  
 DR MGD; MGI:36446; 1gh-4.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003597; Ig\_c1.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00047; Ig1; 2.  
 DR SMART; SM00407; IG1; 2.  
 DR PROSITE; PS50835; IG LIKE; 3.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;  
 KM Alternative splicing; Transmembrane; 3D-structure.  
 FT NON\_TER 1 1  
 FT DOMAIN 1 97 CH1.  
 FT DOMAIN 98 110 HINGE.  
 FT DOMAIN 111 217 CH2.  
 FT DOMAIN 218 324 CH3.  
 FT DISULFID 27 82  
 FT DISULFID 102 102 INTERCHAIN (WITH A LIGHT CHAIN).  
 FT DISULFID 104 104 INTERCHAIN (WITH A HEAVY CHAIN).  
 FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).  
 FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).  
 FT DISULFID 138 198  
 FT CAROHND 174 174 N-LINKED (GLCNAC. . .).  
 FT CAROHND 244 302  
 FT TRANSMEM 340 357  
 FT TRANSMEM 358 393  
 FT DOMAIN 393 393  
 SQ SEQUENCE 393 AA; 43386 MW; 4CC88343B7A1CE27 CRC64;

Query Match 27.5%; Score 940.5; DB 1; Length 393;  
 Best Local Similarity 63.2%; Pred. No. 3,2e-53;  
 Matches 165; Conservative 46; Mismatches 45; Indels 5; Gaps 2;

QY 388 PTWSPVPERKSCDKHTHTPP--CPAPELGSGSVLFFPKPKDTMTMIRTEPVTCVVVDV 445  
 DB 87 PASSSTKVKKTKIVRCGCKPCICTVPEV---SSVFIFPKPKDVTITLTKKVCVVVDI 143  
 QY 446 SHEDEVKFNWYVDDEVHNAKTKPREQYNSTYVWVVLTVLHODMNGKYEKCKVSNK 505  
 DB 144 SKDDEVPQSWVDVEVHTAQTQPREQFNSTFVSVELPLHMODMNGKEPKCRVNSA 203  
 QY 506 ALPAPIETKTSKAKQPREPOVYTLPPSRDELTKNQSILTCIVKGFPSDIAVEMESNGQ 565  
 DB 204 AFPAPIETKTSKTKGRPRAPQVYTLPPRKGWAKDKVSLTCMTDFFPEDITVEMQNNQ 263  
 QY 566 PENNYKTPPVLDSSGSEFLYSKLTVDKSRNQGVEGCSVMHEALNHHYQKSLSPG 625  
 DB 264 PAENYKNTQPTMNTGSIYFVYSKLVNOKSNMEAGMTFTCSVLHBLNHHTEKSLSPG 323  
 QY 626 LQLDETCAADQDELGLMTT 646  
 DB 324 LQLDETCAADQDELGLMTT 344  
 RESULT 21  
 GC\_RABIT  
 ID\_GC\_RABIT STANDARD; PRT; 323 AA.  
 AC P01870;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig gamma chain C region.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OC NCBI\_TaxId=9986;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84030930; PubMed=6313520;  
 RA Bernstein K.E., Alexander C.B., Mage R.G.;  
 RT "Nucleotide sequence of a rabbit IgG heavy chain from the recombinant  
 RT F-1 haplotype.";  
 RL Immunogenetics 18:387-397(1983).  
 RN [2]  
 RP SEQUENCE OF 1-128.  
 RX MEDLINE=76135469; PubMed=1243651;  
 RA Pratt D.M., Mole L.E.;  
 RT "Sequence studies on the constant region of the Fd sections of rabbit  
 RT immunoglobulin G of different allotype.";  
 RL Biochem. J. 151:337-349(1975).  
 RN [3]  
 RP SEQUENCE OF 88-266 FROM N.A.  
 RX MEDLINE=83299917; PubMed=6193512;  
 RA Martens C.L., Moore K.W., Steinmetz M., Hood L., Knight K.L.;  
 RT "Heavy chain genes of rabbit IgG: isolation of a cDNA encoding gamma  
 RT heavy chain and identification of two genomic C gamma genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:6018-6022(1982).  
 RN [4]  
 RP SEQUENCE OF 132-161.  
 RX MEDLINE=70110015; PubMed=5461106;  
 RA Fruchter R.G., Jackson S.A., Mole L.E., Porter R.R.;  
 RT "Sequence studies of the Fd section of the heavy chain of rabbit  
 RT immunoglobulin G.";  
 RL Biochem. J. 116:249-259(1970).  
 RN [5]  
 RP SEQUENCE OF 129-131 AND 155-322.  
 RA Hill R.L., Lebovitz H.E., Fellows R.E. Jr., Delaney R.;  
 RL (In) Killander J. (eds.);  
 RL Gamma globulins, Nobel symp. 3, pp.109-127, Almqvist and Wiksell,  
 RL Stockholm (1967).  
 CC -1- MISCELLANEOUS: Ref.1 sequence has the D12 allotypic marker,  
 CC 104-Thr, and the E14 marker, 185-Thr. Ref.3 has the D11 and E15  
 CC markers and Ref.5 the E15 marker.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1ab-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; M16426; AAA31289.1; -.  
 DR PIR; A91749; GHRB.  
 DR HSSP; P01857; 1PCL.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003597; Ig\_c1.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00047; Ig1; 2.  
 DR SMART; SM00407; IG1; 2.  
 DR PROSITE; PS50835; IG LIKE; 3.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 KM Immunoglobulin domain; Immunoglobulin C region; Repeat.  
 FT NON\_TER 1 1  
 FT DOMAIN 1 96  
 FT DOMAIN 97 213 IG-LIKE 1.  
 FT DOMAIN 214 318 IG-LIKE 2.  
 FT DOMAIN 319 414 IG-LIKE 3.  
 FT VARIANT 104 104 T -> M (IN D11 MARKER).  
 FT VARIANT 185 185 N -> A (IN E15 MARKER).  
 FT CONFLICT 48 48 N -> E (IN REF. 2).  
 FT CONFLICT 71 71 V -> VPV (IN REF. 2).  
 FT CONFLICT 144 144 Q -> E (IN REF. 3 AND 4).  
 FT CONFLICT 173 173 N -> D (IN REF. 5).  
 FT CONFLICT 187 187 Q -> E (IN REF. 3 AND 5).  
 FT CONFLICT 201 201 N -> D (IN REF. 5).  
 FT CONFLICT 218 218 Q -> E (IN REF. 5).  
 FT CONFLICT 233 233 E -> Q (IN REF. 5).  
 FT CONFLICT 246 246 N -> D (IN REF. 5).



```

FT FT CONFLICT 256 256 E -> G (IN REF. 5) .
FT CONFLICT 256 256 N -> D (IN REF. 5) .
FT CONFLICT 256 256 N -> D (IN REF. 5) .
FT CONFLICT 280 280 Y -> W (IN REF. 5) .
FT CONFLICT 284 284 N -> S (IN REF. 5) .
SQ SEQUENCE 323 AA, 35404 MW, 6988AA118D579A8B CRC64;

Query Match 27.0%; Score 922; DB 1; Length 323;
Best Local Similarity 69.3%; Pred. No. 3,8e-52;
Matches 169; Conservative 30; Mismatches 37; Indels 8; Gaps 2

OY 362 SNIKYLPMTSTVEEKSCKHTTCPCPAPELLGGPSVLPFPKPDYLMISRTPEYTV 441
DB 87 TNYTKDKT---VADSTCK---PTCPPELLGGPSVLPFPKPDYLMISRTPEYTV 138
OY 442 VDVSHEDPEVKNYVDGVEVHNAAKTKREEOYNSTYRVSVLTVLHODWLNKGEYCK 501
DB 139 VVDVSGDDPEVQFTWYINNEQRTARPLRBEQFPNSTIIVGSLPTTHQDWLRKGEFKCK 198
OY 502 VSNKALPAPIEKTISKAKGQPRPEQYVTLPSRDELTKQVSLTCLVKGFYPSDIAVEME 561
DB 199 VHNKALPAPIEKTISKAKGQPRPEQYVTLPSRDELTKQVSLTCLVKGFYPSDISEVE 258
OY 562 SNGGQENNYKTPPLVLDSDGSPFLYSKLTVDLSRNOGQGVFCSVWHEALHNHYQKSL 621
DB 259 KNGKADENYKTPPAVLDSGSIYFLNKLSTVPSEWGRGVPTCSVWHEALHNHYQKSL 318
OY 622 LSPG 625
DB 319 RSPG 322

RESULT 22
GCAM_MOUSE STANDARD; PRT; 399 AA.
ID GCAM_MOUSE P01865;
1D 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-2A chain C region, membrane-bound form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83222190; PubMed=6283537;
RA Yamawaki-Kataoka Y., Nakai S., Miyata T., Honjo T.;
RT "Nucleotide sequences of gene segments encoding membrane domains of
RT immunoglobulin gamma chain.";
RT Proc. Natl. Acad. Sci. U.S.A. 79:2623-2627(1982).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Membrane-bound;
CC IsoId=P01865-1; Sequence=Displayed;
CC Name=Secreted;
CC IsoId=P01864-1; Sequence=External;
CC Note=Probably the major isoform;
CC -!- SIMILARITY: Contains 3 immunoglobulin-like domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk.)
CC
DR EMBL, J00471; AAB59661.1; ALT_INIT.
DR PIR, A02154; G2MSAM.
DR PDB, 1KBS; 08-APR-98.
DR PDB, 1YEB; 15-OCT-97.
DR MGD; MGI:96443; Igh-1.

```

DR		InterPro; IPR007110; Ig-like.
DR		InterPro; IPR003597; Ig-cl.
DR		InterPro; IPR003066; Ig_MHC.
DR	Pfam; PF00047; Ig; 2.	
DR	SMART; SM00407; IGcl; 2.	
DR	PROSITE; PS00835; IG_LIKE; 3.	
DR	PROSITE; PS00290; IG_MHC; 1.	
KW	Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;	
KW	Transmembrane; Alternative splicing; 3D-structure; Repeat.	
FT	NON TER	1
FT	DOMAIN	6 98
FT	DOMAIN	121 220
FT	DOMAIN	229 325
FT	DISULFID	15 15
FT	DISULFID	27 82
FT	DISULFID	107 107
FT	DISULFID	110 110
FT	DISULFID	112 112
FT	DISULFID	144 204
FT	DISULFID	250 308
FT	TRANSMEM	346 363
FT	DOMAIN	364 399
FT	CARBOHYD	180 180
SC	SEQUENCE	399 AA; 44020 MM; 4C8138BFAEDDF0 CRC64;
	Query Match	26.8%; Score 915.5; DB 1; Length 399;
	Best Local Similarity	49.6%; Pred. No. 1.3e-51;
	Matches 183; Conservative	48; Mismatches 79; Indels 59; Gaps 7
OY	285 PQALPQVYAGSGNLTALAKTGGKLGHEV-----NLVMRAIQLOKNLICEVGWPSPSKLM	339
DB	34 PEPVYLTVNSGSLSSGVHTFPFAVLQSDLYTLSSVTYVTSSTWPSGITCNVAHPAS----	89
OY	340 LSLKLKENKAQYSKKREKPVWWLNPBAGMWCILLDSGOVLLESNIKVLPTWSTPYEPPKSC	399
DB	90 -----STVDKDIKP-----RG-----PT-----IKP----	106
OY	400 DKHTPCP--CPAPELLGPGSVFLPEPKPKDTLMISRTPEVTCVVADVSHEDPEVKFNMY	457
DB	107 -----CPCKCPAPRLGLGPGSVFIIPPKIKOVLMISTLPITVCVVVDSEDDPDVQISWF	161
OY	458 VDGEVENNAKTKRPREEOYNSTRYRVSVLTVLHODMLNKEKYCKRSNKALPAPIEKITSI	517
DB	162 VNNVEVHAQTQTRHEDYNSTLRVVSALPIOHODMWSKEFKCVNNKDLPAPIERTISK	221
OY	518 AKGPBREQVYTLPSPSRDELTKNOVSLCLKGFYPSPDIAVESNGQPENNYKTTPVL	577
DB	222 PKGSVRAPQVYVLPPEPEEMTKQVTLTCMTDPEDIIYEWATNGKTELNVKTEPVL	281
OY	578 DSDGSPFLYSKLTVDKSRMOOQGNFSSCSVMIEALHNHTOKLSLSPGIQDETCAEAAD	637
DB	282 DSDGSPFYMSKLTAREKKNWVERNSGCVNHGILNHHHTTSSFSRTPGLDLDDVCAEAAD	341
OY	638 GELDGLMTT 646	
DB	342 GELDGLMTT 350	
	RESULT 23	
ID	Gc2_CAVPO	STANDARD; PRT; 329 AA.
AC	P01862;	
DT	21-JUL-1986 (Rel. 01, Created)	
DT	21-JUL-1986 (Rel. 01, Last sequence update)	
DT	15-JUL-1999 (Rel. 38, Last annotation update)	
DE	Ig gamma-2 chain C region.	
OS	Cavia porcellus (Guinea pig).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
NCBI_TaxId=10141;	Mammalia; Eutheria; Rodentia; Hystricognathi; Cavidae; Cavia.	
RP	Tiltschmann T.M.;	



```

RL Submitted (APR-1975) to the PIR data bank.
RN [2]
RP SEQUENCE OF 4-68.
RX MEDLINE=71058471; PubMed=5538606;
RA Birhnein B.K., Huseaini Q.Z., Gebra J.J.;
RT "Structure of heavy chain from strain 13 guinea pig
  immunoglobulin-G(12). 3. Amino acid sequence of the
  half-cysteine joining heavy and light chains.";
RL Biochemistry 10:18-25(1971).
RN [3]
RP SEQUENCE OF 69-133 AND 312-329.
RX MEDLINE=71058486; PubMed=5538616;
RA Turner K.J., Gebra J.J.;
RT "Structure of heavy chain from strain 13 guinea pig
  immunoglobulin-G(12). II. Amino acid sequence of the
  and hinge region cyanoogen bromide fragments.";
RL Biochemistry 10:9-17(1971).
RN [4]
RP SEQUENCE OF 134-226.
RX MEDLINE=75036072; PubMed=4429665;
RA Tracey D.E., Gebra J.J.;
RT "Primary structure of the CH2 homology region from guinea pig IgG2
  antibodies.";
RL Biochemistry 13:4796-4803(1974).
RN [5]
RP SEQUENCE OF 227-311.
RX MEDLINE=75036073; PubMed=4609467;
RA Trischmann T.M., Gebra J.J.;
RT "Primary structure of the CH3 homology region from guinea pig IgG2
  antibodies.";
RL Biochemistry 13:4804-4811(1974).
RN [6]
RP DISULFIDE BONDS.
RX MEDLINE=71058474; PubMed=4922544;
RA Oliveira B., Lamm M.E.;
RT "Interchain disulfide bridges of guinea pig gamma-2-immunoglobulin.",
  Biochemistry 10:26-31(1971).
RL -1- MISCELLANEOUS: This chain was isolated from pooled serum of strain
  13 inbred guinea pigs.
CC PIR: A94553; G2GP.
DR HSSP; P01842; 77AB.
DR InterPro; IPR007110; Ig-1-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; glycoprotein.
FT NON_TER 1 1
FT DISULFID 16 16 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 28 79
FT DISULFID 105 105 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 110 110 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 142 202
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .).
FT DISULFID 248 308
SQ SEQUENCE 329 AA; 36074 MW; 5D231B7164D1FB9A CXC64;
Query Match 26.6%; Score 906.5; DB 1; Length 329;
Best Local Similarity 58.1%; Pred. No. 3.8e-51;
Matches 176; Conservative 31; Mismatches 57; Indels 39; Gaps 4;
QY 325 NLTCGVGPTSPKMLSLKLNKAKVSKREKPVVNLPEAGMOCCLISDGYLLSNT 384
DB 64 SLTGVVTVPSOQKATCNVAHPASSRTKVKTEVPI----- 97
QY 385 KVLPTWSTVPPPKSCDKTHTCPAPRLGSPVFLFPKPKDTMTSRTPEYTCVVD 444
DB 98 -----RTP-ZRPBC---TCKPCDPNLDGSPSVFLFPKPKDTMTSLTRVYTCVVD 146
QY 445 VSHEDPEVKFMVYDGVENHNAKTKPREBOYNSTVRVSVLTVLHODWLNGEKYCKVSN 504

```

```

DB 147 VSODEPEVQFTWFEVDNKPVNAAETKPEVEQYNTTFRVESVLPLOHODWLNKGEKFCVYN 206
QY 505 KALPAPIEKTISAKAQGPREFROYTLPPSRDELTJKNVSLTCLVKGYPDIWEMSSNG 564
DB 207 KALPAPIEKTISKKGAPRPMDVTLPPSRDELSKRSVYTCILINFPPADHVMASNR 266
QY 565 QP--ENNYKTPPVLDSDGSFPLYSKLTVDKSRMOQSNVSCSYMEBALNHYTKSLSL 622
DB 267 VPVSEKEYKNTPTPIEDADSGSYFLYSKLTVDKSNAMDQCTVTCGVMHEALNHYTKSL 326
QY 623 SPG 625
DB 327 SPG 329
RESULT 24
GCBM_MOUSE STANDARD; PRT; 405 AA.
ID GCBM_MOUSE
AC P01867;
DT 21-JUL-1986 (Ref. 01, Created)
DT 01-AUG-1991 (Ref. 19, Last sequence update)
DT 15-MAR-2004 (Ref. 43, Last annotation update)
DE Ig gamma-2B chain C region, membrane-bound form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 335-405 FROM N.A.
RX MEDLINE=82221190; PubMed=6283537;
RA Yamawaki-Kataoka Y., Nakai S., Miyata T.;
RT "Nucleotide sequences of gene segments encoding membrane domains of
  immunoglobulin gamma chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:2623-2627(1982).
RN [2]
RP SEQUENCE OF 335-378 FROM N.A.
RX MEDLINE=82115295; PubMed=6799207;
RA Rogers J., Choi E., Souza L., Carter C., Word C.J., Kuehl M.,
RA Eisenberg D., Wall R.;
RT "Gene segments encoding transmembrane carboxyl termini of
  immunoglobulin gamma chains.";
RL Cell 26:19-27(1981).
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Membrane-bound;
CC IsoId=P01867-1; Sequence=Displayed;
CC Name=Secreted;
CC IsoId=P01866-1; Sequence=External;
CC Note=May be the major isoform;
CC -1- PM: DISULFIDE BONDS BY HOMOLOG WITH OTHER IG GAMMA CHAINS.
CC -1- MISCELLANEOUS: The sequence of residues 1-335 is assumed to be
  identical with the corresponding region of the secreted form.
CC -1- MISCELLANEOUS: The a allele sequence is shown.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.ebi.ac.uk/announce/
  or send an email to license@ebi.ac.uk).
CC -----
DB EMBL; J00462; AAB59659.1; ALT_INIT.
DR PIR; C02154; GZMSM.
DR PDB; 1GIC; 11-MAR-03.
DR MGD; MGI:96445; Igh-3.
DR InterPro; IPR007110; Ig-1-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IGc1; 2.

```



```

DR PROSITE; PS50835; IG_LIKE; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Transmembrane;
KW Alternative splicing; 3d-structure; Repeat.
FT NON_TER 1
FT DOMAIN 6 98 IG-LIKE 1.
FT DOMAIN 127 226 IG-LIKE 2.
FT DOMAIN 235 331 IG-LIKE 3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 82 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 115 115 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 118 118 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 150 210 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 256 314
FT TRANSMEM 352 369
FT DOMAIN 370
SQ SEQUENCE 405 AA; 44330 MW; 89B3CF0A9B6D49FA CRC64;

Query Match
Best Local Similarity 47.4%; Score 884.5; DB 1; Length 405;
Matches 174; Conservative 53; Mismatches 91; Indels 49; Gaps 4;

QY 285 POLAPQYAGSGNLTALAE-----KTGKLHDEVNLVVMRATOLQKLTCEVWGPTSPKLM 339
DB 34 PPSVITVNMSSGSSVHTFPALLOSGLYTMSSSVTPSPQVITCSVAHPAS-STT 92
QY 340 LSLKLENKAAVSKREKPVVNLPEAGWQCLISDSGQVLLSNIKVLPTWSTPVEPKSC 399
DB 93 YRKLE-----PSPGPISTNP-----C 109
QY 400 DKTHNCPAPABELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKRWYD 459
DB 110 PPKCKCHKPAPRLGEGSPVFLFPPIKQVLMISLTPKVTGVVDSEDDPVQISWFPN 169
QY 460 GVEVNAKTKPREEQNSTYRVVSVLTJLHODMLNKEKCYKSNKALPAPIEKTISKAK 519
DB 170 NVEVHTAQVQTHREDNSTIRVSVTLPIQHODMSKEKCKVNNKDLSPIERITISKK 229
QY 520 GGPREFQVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIAVWESNQGPENNYKTPPVLD 579
DB 230 GLVRAQVYILPPPAEQLSRKDVSLTCLVGVFPGDISVEMTNGHTEENYKDTAPVLD 289
QY 580 DGSFPLSLTLYDKSKWQGNVFGSGVMEALHNHTOKSLSLSPLOLDEFCARQDGE 639
DB 290 DGSYFLYSLLNMTKTSKMEKTDSPSCVNRHGLKNYYLKXTISRSPGLDLDDICAEAKDGE 349
QY 640 LDGLWTT 646
DB 350 LDGLWTT 356

RESULT 25
GC3_MOUSE STANDARD; PRT; 329 AA.
AC P22536;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE IG gamma-3 chain C region, secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85027161; PubMed=6092053;
RA Mele J.A., Word C.J., Rimm D., Der-Balan G.P., Martinez H.M.,
RA Tucker P.W., Blattner F.R.;
RT "Structural analysis of the murine IgG3 constant region gene.";
RL EMBO J. 3:2041-2046(1984).
-----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; J00451; -; NOT_ANNOTATED_CDS.
DR F01; B02156; G3MSC.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Transmembrane; Alternative splicing.
FT NON_TER 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 113 HINGE.
FT DOMAIN 114 223 CH2.
FT DOMAIN 224 327 CH3.
SQ SEQUENCE 329 AA; 36228 MW; P45827174182BAD6 CRC64;

Query Match
Best Local Similarity 58.1%; Score 849.5; DB 1; Length 329;
Matches 154; Conservative 42; Mismatches 48; Indels 21; Gaps 2;

QY 382 SNIKVLPMTWSTPVEPKSCDKTH-----TCPP-----CPABELLGGSPVF 420
DB 64 SSVTVPSSTPSPQVINCVAHPASKTELIRIEPRIPKSTPSSGCPGNILGSPSVF 123
QY 421 LFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKRWYDGVVNAKTKPREEQNSTYR 480
DB 124 IFPPKPKALMISLTPKVTGVVDSEDDPVHVSFMDNKEVHTAMQPREAQVNSTFR 183
QY 481 VSVSLTJLHODMLNKEKCYKSNKALPAPIEKTISKAKGPREQVYTLPPSRDELTKN 540
DB 184 VVSALPIQHODMWGMEKCKVNNKALPAPIERTITSKPKGAQVQVYTIIPPREQMSKK 243
QY 541 QVSLTCLVKGFPYPSDIAVWESNQGPENNYKTPPVLDSDSFLYSLLTLYDKSWQGN 600
DB 244 KVSLLTCLVTFPSEASIVEMERNGELQDYKNTPEILDSDYFLYSLLTLYDTSWLDGE 303
QY 601 VFGSGVMEALHNHTOKSLSPG 625
DB 304 IFTGCVVMEALHNHTOKLSRSPG 328

RESULT 26
GCB_RAT STANDARD; PRT; 333 AA.
AC P20761;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE IG gamma-2B chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89232738; PubMed=3149946;
RA Brueggemann M.;
RT "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
RL Gene 74:473-482(1988).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
DR PIR; PS0018; PS0018.
DR HSSP; P01842; 7FAB.
-----

```



```

DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF000407; Ig; 3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS00835; IG-LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
DR Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON TER 1 1
FT DOMAIN 6 96 IG-LIKE 1.
FT DOMAIN 124 223 IG-LIKE 2.
FT DOMAIN 232 328 IG-LIKE 3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 80 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 115 115 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 147 207
FT DISULFID 253 311
SQ SEQUENCE 333 AA; 36497 MW; 55F8B64D48D460A6 CRC64;

Query Match 24.7%; Score 844; DB 1; Length 333;
Best Local Similarity 61.7%; Pred. No. 3.9e-47;
Matches 153; Conservative 34; Mismatches 51; Indels 10; Gaps 1;

OY PWSSTPVPK-----SCDKTHTCPCPAPELGSPSVFLPPPKDXTLMISRTPE 437
DB 85 PASSRKVKYKVRNRNGIGHKCPTCTCHKCPVPELLGGPSVFIFFPKPKIILISQNAK 144
OY 438 VTCVVVDVSHEDPEVKFMYVDGVEVHNAAKTRPEEQYNSITVRVSVLTVLHQMUNGKE 497
DB 145 VTCVVVDVSEEDPDVQFPMFNVNVEVHTAQTQPREQYNSTFRVVSALPIQHQMWSGKE 204
OY 498 YKCKSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIA 557
DB 205 FKCKRNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIA 264
OY 558 VEMESNGQPNENYKTPPVLDSDGSFPLYSKLTVDKSRMOCGVNFCSEVMEHALHNHYTQ 617
DB 265 VEMISNGIIEKYNKTEPVMDSGSFMYSKLVNFRSNWRAPVCSVVAHEGLHNHVE 324
OY 618 KSLSLSPG 625
DB 325 KSLSLSPG 332

RESULT 27
GCL_RAT STANDARD; PRT; 326 AA.
ID_GCL_RAT
AC P20759;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig gamma-1 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89232738; PubMed=3149946;
RA Brueggemann M.;
RL "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
RL Gene 74:473-482(1988).
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF000407; Ig; 3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS00835; IG-LIKE; 3.

```

```

DR PROSITE; PS00290; IG_MHC; 1.
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON TER 1 1
FT DOMAIN 98 97 CH1.
FT DOMAIN 113 219 HINGE.
FT DOMAIN 220 326 CH2.
FT DISULFID 27 82 CH3.
FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 111 111 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 140 200
FT CARBOHYD 176 304
FT CARBOHYD 176 176
SQ SEQUENCE 326 AA; 35946 MW; 013BAB45EFA9B9DA CRC64;

Query Match 24.3%; Score 828; DB 1; Length 326;
Best Local Similarity 55.1%; Pred. No. 4.1e-46;
Matches 150; Conservative 48; Mismatches 48; Indels 26; Gaps 4;

OY 375 SGCVLLESNIKVLPTWSTPVPKSCDKTH-----TCPCPAPELG 416
DB 59 SGLYTLTSSVTV-PSSTWPSQVTVCNVAHPASSTKDKKIVPRNCGDCKPC---ICTG 113
OY 417 ---PSVFLPPPKKPDITLMISRTPEVTCVVVDVSHEDPEVKFMYVDGVEVHNAAKTRPEE 473
DB 114 SEVSVVIFPPKPKDVLITLTPKVTICVVVDISQDDPEVAFSNVDDVEVHTAQTREPE 173
OY 474 QNSTYRVSVLTVLHQMUNGKEVYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPS 533
DB 174 QFNSTFRSVSELDILHQMUNGRTFRCKVTSAAFPSPIEKTIISKPEGRTPVPHYTWSP 233
OY 534 RDELTKNQVSLTCLVKGFYPSDIAVEMESNGQPNENYKTPPVLDSDGSFPLYSKLTVDK 593
DB 234 KEETQNEVSIITCVKGFYPPDIYVEMQNGQDENYKNTPTWDTGSIPLYSKLVNKK 293
OY 594 SRMOCGVNFCSEVMEHALHNHYTQKSLSPG 625
DB 294 EKMOGNTFTCSVLHDEGLHNHTEKSLSPG 325

RESULT 28
GCL_MOUSE STANDARD; PRT; 324 AA.
ID_GCL_MOUSE
AC P01868;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig gamma-1 chain C region secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80045036; PubMed=115593;
RA Honjo T.; Obara M.; Yamawaki-Kataoka Y.; Kataoka T.; Kawakami T.;
RA Takahashi N.; Mano Y.;
RT "Cloning and complete nucleotide sequence of mouse immunoglobulin
RT gamma 1 chain gene.";
RL Cell 18:559-568(1979).
RN [2]
RP SEQUENCE OF 76-324 FROM N.A. (MYELOMA PROTEIN MOPC 31C).
RX MEDLINE=80202553; PubMed=6769752;
RA Obara M.; Yamawaki-Kataoka Y.; Takahashi N.; Kataoka T.; Shimizu A.;
RA Mano Y.; Seidman J.G.; Peterlin B.M.; Leder P.; Honjo T.;
RT "Immunoglobulin gamma 1 heavy chain gene: structural gene sequences
RT cloned in a bacterial plasmid.";
RL Gene 9:87-97(1980).
RN [3]
RP SEQUENCE OF 70-322 FROM N.A. (MYELOMA PROTEIN MOPC 21).
RX MEDLINE=80012837; PubMed=113776;

```







```

OY 375 SGQVLLSNIKV-LFTWSTPVEPKSCDKTHTC-----PP---CPAEL 413
DB 59 SGLVLTSSSVTPSSTWSS--QTVCSVAHPATKSNLIKRIEPRPKRPFDICSCDN 116
OY 414 LGSPSVFLPPPKKDTLMISRTPEYTCVVVDVSHEDPEVKFMWYVDGVEYNATKPKREE 473
DB 117 LGRPSVFLPPPKKDTLMITLTPKTCVVVDVSEEPVQPSWFDVNRVFTAQTPHEE 176
OY 474 QYNSTYRVSVLTVLHQMPLNGKEKCYKNALPAPIEKTISKAKGQPREPQVYTLPPS 533
DB 177 QLNQFRVYVSTLTIHQDMGSKERKCKVNNKDLPSLEKTIISKRGAKTPOVYITPP 236
OY 534 RDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLDSDGSFFLYSKLTVDK 593
DB 237 REQSKNKVSLTCWTSFYFPASISVEMERNGELBDYKNTLPVLDSDESYFLYSKLSYDT 296
OY 594 SRMOGNVFCSGVMEALHNHYTKSLSPG 625
DB 297 DSMKGDIVTCSVMEALHNHTQKLSRSPG 328

RESULT 30
GCAB_MOUSE STANDARD; PRT; 335 AA.
AC P01864;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig gamma-2a chain C region secreted form (B allele).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA MEDLINE=82037861; PubMed=6170065;
RX Schreier P.H., Bochtell A.L.M., Mueller-Hill B., Baltimore D.;
RT "Multiple differences between the nucleic acid sequences of the
IG gamma and IG2ab alleles of the mouse.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:4495-4499(1981).
RN [2]
RP SEQUENCE.
RA MEDLINE=82037777; PubMed=6794027;
RX Dognin M.J., Lauwereys M., Strosberg A.D.;
RT "Multiple amino acid substitutions between murine gamma 2a heavy
chain Fc regions of Ig1a and Ig1b allotypic forms.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:4031-4035(1981).
CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Secreted;
CC IsoId=P01864-1; Sequence=Displayed;
CC Note=Probably the major isoform;
CC Name=Membrane-bound;
CC IsoId=P01865-1; Sequence=External;
CC -1- MISCELLANEOUS: The sequence differs from that of the a allele,
from BALB/c mice, at 15 of the positions.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; J00479; -; NOT_ANNOTATED_CDS.
CC PIR; A02153; G2MSAB.
CC PDB; 1BOG; 23-MAR-99.
CC PDB; 1HH6; 26-JAN-01.

```

```

DR PDB; 1HH9; 24-JUL-03.
DR PDB; 1HH6; 08-FEB-01.
DR InterPro; IPR007110; Ig_1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003066; Ig_MHC.
DR Pfam; PF00047; Ig_2.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE_3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Alternative splicing;
KW 3D-structure; Repeat.
FT NON TER 1
FT DOMAIN 6 98 IG-LIKE 1.
FT DOMAIN 126 225 IG-LIKE 2.
FT DOMAIN 234 330 IG-LIKE 3.
SQ SEQUENCE 335 AA; 36596 MW; FA3382792CBB136C CRC64;

Query Match 23.9%; Score 816; DB 1; Length 335;
Best Local Similarity 45.7%; Pred. No. 2,5e-45;
Matches 158; Conservative 52; Mismatches 86; Indels 50; Gaps 3;

OY 285 POALPOVAGSGNLTIALEA-----KTGKLGHEVNLVVMRATQLOKNTCEVWGPTSPKLM 339
DB 34 PEVYTLTWNSSGLSSGVHTPRLALQSGLYTLSSVYTSNTWPSQITCNVHPAS----- 89
OY 340 LSLKENKEAKVSKREKPVVNLNPEAGMQCLLSDSQVLLSNIKVLPTWSTPVEPKSC 399
DB 90 -----SKRVKRIEP-----RVPTQNPC 108
OY 400 DKHTCPCPAPBELGSPVFLPPPKKDTLMISRTPEYTCVVVDVSHEDPEVKFMWYVD 459
DB 109 PPHQRPVCAAPDLGLGSPVFIPEPKIKVLMISLSMVCVVVDSEDDVDQISMFVN 168
OY 460 GVEVHNKATKPREQVSTYRVSVLTVLHQMPLNGKEKCYKNALPAPIEKTISKAK 519
DB 169 NVEVHTAQTOHREDYNTLRVVSALPIHQDMGSGEFPCKNNRRLPSPIETIKRPR 228
OY 520 GQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTPPVLD 579
DB 229 GPVAPQVYVLPPEAEWTKKERSLTGMINGFLPAEIAVDMWTSNGRTQVYKNTATVLD 286
OY 580 DGSFELYSKLTVDKSRMOGNVFCSGVMEALHNHYTKSLSPG 625
DB 289 DGSYFMYSKLRVOKSTWERSGLPACSVMEALHNHTTKTISRIG 334

RESULT 31
GCCA_MOUSE STANDARD; PRT; 330 AA.
AC P01863;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-2a chain C region, A allele.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=81076554; PubMed=6777755;
RX Sikorav J.-L., Auffray C., Rougeon F.;
RT "Structure of the constant and 3' untranslated regions of the murine
Balb/c gamma 2a heavy chain messenger RNA.";
RL Nucleic Acids Res. 8:3143-3155(1980).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=81198976; PubMed=6262729;
RX Yamawaki-Kataoka Y., Miyata T., Honjo T.;
RT "The complete nucleotide sequence of mouse immunoglobulin gamma 2a gene
and evolution of heavy chain genes: further evidence for intervening
sequence-mediated domain transfer.";
RL Nucleic Acids Res. 9:1365-1381(1981).

```



```

RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=81223894; PubMed=6787604;
RA Olio R., Aufrey C., Mochamps C., Rougeon F.;
RT "Comparison of mouse immunoglobulin gamma 2a and gamma 2b chain genes
RT suggests that exons can be exchanged between genes in a multigenic
RT family.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:2442-2446 (1981).
RN [4]
RP MYELOMA PROTEIN MOPC 173.
RX MEDLINE=74175517; PubMed=4831970;
RA Bougouie A., Fougereau M., Rocca-Serra J.;
RT "Determination of the primary structure of a mouse IgG2a
RT immunoglobulin: amino-acid sequence of the Fc fragment. Implications
RT for the evolution of immunoglobulin structure and function.";
RL Eur. J. Biochem. 43:423-435 (1974).
RN [5]
RP DISULFIDE BONDS.
RX MEDLINE=73056887; PubMed=4565406;
RA de Preval C., Fougereau M.;
RT "Determination of the primary structure of a mouse gamma G2a
RL immunoglobulin. Identification of the disulfide bridges.";
RL Eur. J. Biochem. 30:452-462 (1972).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; V00798; CAA4178.1; -
DR PIR; A02152; G2MSA.
DR PDB; 1E4W; 12-JUL-01.
DR PDB; 1E4X; 12-JUL-01.
DR PDB; 1MWU; 06-MAY-99.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003587; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS00835; IG LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
DR Immunoglobulin domain; Immunoglobulin C region; 3D-structure; Repeat.
DR NON_TER 1
FT DOMAIN 1 98 IG-LIKE 1.
FT DOMAIN 121 220 IG-LIKE 2.
FT DOMAIN 229 335 IG-LIKE 3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 82 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 110 110 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 144 204 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 250 308
FT MOD_RES 330 330 REMOVED POST-TRANSLATIONALLY.
FT SEQUENCE 330 AA; 36389 MW; B84361C5445A864 CRC64;
SQ
Query Match 23.9%; Score 814.5; DB 1; Length 330;
Best Local Similarity 47.4%; Pred. No. 36-45; Indels 59; Gaps 7;
Matches 165; Conservative 47; Mismatches 77;

```

```

DB 107 -----CPCKCAPRLTGGPSVFIFRPKIKQVLMISLSPITCVVVDVSEDDPDVQISMF 161
QY 458 VDGVEVHNAKTKPREEQNSTYRVSVLTVLHODMLNGEKYKCVSNKALPAPIEKTISK 517
DB 162 VNNVEVHTAQOTGTHBEDYVSTLRVVSALPIQHODWMSGKEFCYKNNKDLPAPIERTISK 221
QY 518 AKGQREPOVYVLPSPRSRELTQNOVSLTCIVKGFPSDIAVEMENSGQPENNYKTPPTVL 577
DB 222 PKGSVRAPOVYVLPPEBEEMTKQVTLTCMVTFDEPEDIYVEMTNGKTELNVKTEPVL 281
QY 578 DSDGSFFLYSKLTVDSRWQGNVSCSMHEALNNHTOKSLSPG 625
DB 282 DSDGSYFMYSKLRVEKKNWVERNSYSCSVHBEGLHNHTTISFSRTPG 329
RESULT 32
GCA_RAT
ID GCA_RAT STANDARD: PRT; 322 AA.
AC P20760;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-2a chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89232738; PubMed=3149946;
RA Bruggemann M.;
RT "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
RL Gene 74:473-482 (1988).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M13804; AAA41376.1; ALT_INIT.
DR PIR; PS0019; PS0019.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS00835; IG LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
DR Immunoglobulin domain; Immunoglobulin C region; Repeat.
DR NON_TER 1
FT DOMAIN 1 98 IG-LIKE 1.
FT DOMAIN 115 212 IG-LIKE 2.
FT DOMAIN 221 317 IG-LIKE 3.
FT DISULFID 27 82 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 105 105 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 136 196 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 242 300
FT CARBOHYD 172 172
FT SEQUENCE 322 AA; 35186 MW; BEBA136A9D801EDB CRC64;
SQ
Query Match 23.3%; Score 795; DB 1; Length 322;
Best Local Similarity 56.4%; Pred. No. 5-26-44;
Matches 150; Conservative 39; Mismatches 59; Indels 18; Gaps 5;

```



```

Db      59 SGLYLTSSVTPSPSTWSSQAVTCNVAPASSTKVCKIVPRECNPCCGCTGSEV---SSV 115
Oy      420 FLFPKPKMDTLMISTRPVTCVVVDVSHEDPEVKFNNVVDVGVHNKTRKEQVNSTY 479
Db      116 FLFPKPKMDVLTTLTPKVTCTVVDVSDVSHEDPEVKFNNVVDVGVHNKTRKEQVNSTY 175
Oy      480 RVSVLTALVHODMNGKKEKCKVSKALPAPIEKTISKAKQPREPVYTLPPSRDELTK 539
Db      176 RVSELPPIVHRDWMNGKTFCKCKVSGAPPAIEKSIKSESTPGPGVYTTAPKREMTQ 235
Oy      540 NQVSLTCLVKGFPYSDIAVEMESNGQPENNKTTTPVLDSGSPFLSKLTVDKSRWQG 599
Db      236 SQVSLTCVWKGFPYSDIATEWKMNQOPENNKTTPTMTDTSYFLYSKLVNKKETWQOG 295
Oy      600 NVFSGSVMEALHNHYTKSLSPG 625
Db      296 NTFCTSVLHGBLHNHTEKSLSHSPG 321

RESULT 33
GCB_MOUSE STANDARD; PRT; 336 AA.
AC P01866;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig gamma-2b chain C region secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ALLELE A).
RX MEDLINE=80120716; Pubmed=6766534;
RA Yamawaki-Kataoka Y., Kataoka T., Takahashi N., Obata M., Honjo T.;
RT "Complete nucleotide sequence of immunoglobulin gamma2b chain gene
RL cloned from newborn mouse DNA.";
RL Nucleic Acids Res. 1980;8:786-789(1980).
RN [2]
RP SEQUENCE FROM N.A. (MPC 11).
RX MEDLINE=80081501; Pubmed=117548;
RA Tucker P.W., Marcu K.B., Slightom J.L., Blattner F.R.;
RT "Structure of the constant and 3' untranslated regions of the murine
RL gamma 2b heavy chain messenger RNA.";
RL Science 206:1299-1303(1979).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=80081502; Pubmed=117549;
RA Tucker P.W., Marcu K.B., Newell N., Richards J., Blattner F.R.;
RT "Sequence of the cloned gene for the constant region of murine gamma
RL 2b immunoglobulin heavy chain.";
RL Science 206:1303-1306(1979).
RN [4]
RP SEQUENCE FROM N.A. (ALLELE B).
RX MEDLINE=82173203; Pubmed=6803173;
RA Ojio R., Rougeon F.;
RT "Mouse immunoglobulin allotypes: post-duplication divergence of gamma
RL 2a and gamma 2b chain genes.";
RL Nature 296:761-763(1982).
RN [5]
RP CARBOHYDRATE-LINKAGE SITE THR-105.
RX MEDLINE=94216359; Pubmed=7512967;
RA Kim H., Yamaguchi Y., Masuda K., Matsumura C., Yamamoto K.,
RA Irimura T., Takahashi N., Kato K., Arata Y.;
RT "O-glycosylation in hinge region of mouse immunoglobulin G2b.";
RL J. Biol. Chem. 269:12345-12350(1994).
CC -1- SUBCELLULAR LOCATION: Secreted (potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Secreted;
CC IsoId=P01866-1; Sequence=Displayed;
CC Note=May be the major isoform;

```

---

```

CC CC Name=Membrane-bound;
CC CC IsoId=P01867-1; Sequence=External;
CC -1- PTM: O-LINKED GLYCAN CONSISTS OF GAL-GALNAC DISACCHARIDE WHICH IS
CC MODIFIED WITH 2 STATIC ACID RESIDUES.
CC -1- PTM: DISULFIDE BONDS BY HOMOLOG WITH OTHER IG GAMMA CHAINS.
CC -1- MISCELLANEOUS: The a allele sequence is shown.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
DR PIR: S25057; G2MS11.
DR HSSP: P01842; 7FAB.
DR InterPro: IPR007110; IG-Like.
DR InterPro: IPR003597; IG cl.
DR InterPro: IPR003066; IG_MHC.
DR Pfam: PF00047; Ig; 3.
DR SMART: SM00407; IgC1; 2.
DR PROSITE: PS50835; IG_LIKE; 3.
DR PROSITE: PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing; Repeat.
FT NON_TER 1 1
FT DOMAIN 6 98 IG-LIKE 1.
FT DOMAIN 127 226 IG-LIKE 2.
FT DOMAIN 235 331 IG-LIKE 3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 82 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 115 115 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 118 118 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 150 210
FT DISULFID 256 314
FT CARBOHYD 105 105
FT MOD_RES 336 336
FT VARIANT 163 163 Q -> R (IN ALLELE B).
FT VARIANT 194 194 T -> A (IN ALLELE B).
FT VARIANT 300 300 M -> D (IN ALLELE B).
FT VARIANT 301 301 M -> I (IN ALLELE B).
FT CONFLICT 25 25 L -> S (IN REF. 2 AND 3).
FT CONFLICT 36 36 S -> P (IN REF. 2 AND 3).
FT CONFLICT 239 239 I -> T (IN REF. 2 AND 3).
SQ SEQUENCE 336 AA; 36658 MW; 7D879662607C356E CRC64;

Query Match 23.1%; Score 788.5; DB 1; Length 336;
Best Local Similarly 45.4%; Pred. No. 1,4e-43;
Matches 157; Conservative 51; Mismatches 89; Indels 49; Gaps 4;

Oy 285 POLAPQVAGSGNLTALAE-----KTGKLGHEVNLVYMRATQLOKNTCEVWGPTSPKLM 339
Db 34 PESVTVTWNSGSLSSSVHTPPALLQSGLYTMSSSVTPSSITWPSQIVTCVAPAS-STT 92
Oy 340 LSLLENKEAKVSKREKPVVWLNPEAGMOCLLDSQGVLLSNIKVLPMTSTPEBPKC 399
Db 93 VDKKLE-----PSGPSTINP-----C 109

Oy 400 DKHTPCPCAPPELLGSPVFLFPKPKDLMISRTPEVTCVVVDVSHEDPEVKFNNVVD 459
Db 110 PPKCKCHKCPAPNIEGSPVFIPEPNIKDLMISRTPEVTCVVVDVSHEDPEVKFNNVVD 169
Oy 460 GVEVHNKAKTPREQVSTRVVSVTLVTLHODMNGKKEKCKVSKALPAPIEKTISKAK 519
Db 170 NVEVHTAQOTHRDYNSTIRVSVTLPIQHODMNGKKEKCKVSKALPAPIEKTISKAK 229

Oy 520 GQPREPQVYTLPPSRDELTKNOVSLTCLVKGFPYSDIAVEMESNGQPENNKTTTPVLDS 579
Db 230 GLVAPAPQVYTLPPSRDELTKNOVSLTCLVKGFPYSDIAVEMESNGQPENNKTTTPVLDS 289
Oy 580 DGSFELYSLKLTVDKSRWQGNVFSQVMEALHNHYTKSLSPG 625
Db 290 DGSFYISKLMNMTSKWEKTDTSFSCNVRHGLKLVYIKTISRSPG 335

RESULT 34
MUCM_MOUSE STANDARD; PRT; 476 AA.

```



AC P01873;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig mu chain C region membrane-bound form.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE OF 433-476 FROM N.A.  
 RX MEDLINE=80222874; PubMed=6771020;  
 RA Early P., Rogers J., Davis M., Calame K., Bond M., Wall R.,  
 RA Hood L.;  
 RT "Two mRNAs can be produced from a single immunoglobulin mu gene by  
 RT alternative RNA processing pathways";  
 RL Cell 20:313-319 (1980).  
 RN [2]  
 RP SEQUENCE OF 410-476 FROM N.A. (MYELOMA MOPC 104E).  
 RX MEDLINE=80222873; PubMed=6771019;  
 RA Rogers J., Early P., Carter C., Calame K., Bond M., Hood L.,  
 RA Wall R.;  
 RT "Two mRNAs with different 3' ends encode membrane-bound and secreted  
 RT forms of immunoglobulin mu chain";  
 RL Cell 20:303-312 (1980).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Comment=During differentiation, B lymphocytes switch from  
 CC expression of isoform Membrane-bound to isoform Secreted;  
 CC Name=Membrane-bound;  
 CC IsoId=P01873-1; Sequence=Displayed;  
 CC Name=Secreted;  
 CC IsoId=P01872-1; Sequence=External;  
 CC -1- MISCELLANEOUS: The sequence of residues 1-409 is assumed to be  
 CC identical with the corresponding region of the secreted form.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL, V00821; CAA24202.1; -.  
 DR PIR; A02167; MHSM.  
 DR HSSP; P01857; 1FC1.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003597; IG\_c1.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR Pfam; PF00047; Ig; 4.  
 DR SMART; SM00407; IgC1; 2.  
 DR PROSITE; PS00835; IG\_1like; 4.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;  
 KW Alternative splicing; Transmembrane.  
 FT NON\_TER 1 1  
 FT DOMAIN 1 105  
 FT 217 CH1.  
 FT DOMAIN 106 217  
 FT 218 CH2.  
 FT DOMAIN 218 324  
 FT 325 CH3.  
 FT 436 CH4.  
 FT TRANSMEM 456 473  
 FT 473 POTENTIAL.  
 FT DISULFID 14 14  
 FT 89 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).  
 FT DISULFID 28 89  
 FT 199 BY SIMILARITY.  
 FT DISULFID 136 199  
 FT 216 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).  
 FT DISULFID 246 305  
 FT 305 BY SIMILARITY.  
 FT DISULFID 293 293  
 FT 415 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).  
 FT DISULFID 353 415  
 FT 46 BY SIMILARITY.  
 FT CARBOHYD 46 46  
 FT 211 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 211 211  
 FT 243 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 243 243  
 FT 261 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 261 261

SEQ SEQUENCE 476 AA; 52656 MW; 8D476575A5204071 CRC64;  
 Query Match 11.0%; Score 375.5; DB 1; Length 476;  
 Best Local Similarity 22.9%; Pred. No. 6.8e-17;  
 Matches 127; Conservative 83; Mismatches 188; Indels 157; Gaps 18;  
 QY 144 ESPPSSSSVQCRSP-RGKNIQGGKTLSSVSQLDELQDSGTWCTVLQNKQKVEFKIDIVVL 202  
 DB 3 QSPFVFPVLGCEPSLSDKNLVAMGCLARDLPSTISFTM----- 43  
 QY 203 AFQKASTIVYKKEGQVEFPFLAFTVEKLTGSGELMWQARASSSKSMTFPLDKNKVS 262  
 DB 44 -----YQNTTEVIO-----GIRTFPLRTGCKXVA 68  
 QY 263 VKRYVDPKLQW-----GKKLPLHLTLQALPOYAGSSGLTLALEAKTKYL 308  
 DB 69 TSQVLSPKSLIEGSDYLVCKIKHYGKNRDLHVPF-----AVAM 110  
 QY 309 HOEVLAVY-----MRATOLQNLTCVWGFTSPFLMLSLKLENKAQVSRKRPVYVLN 362  
 DB 111 NPNNVVFPPRDGSPGAPRKSKLICEATNTTPKRTYSWLKDGKLVESGFTTDPVTLEN 170  
 QY 363 ----PEA-----GMQCLLSDSQVLLIESNLIKVLPTWSTPVEPKSCDK 401  
 DB 171 KGSFQTKVAVISTLTISEIDMLNLNVYTCRVYDHRGLTFK-----NV 212  
 QY 402 THTCPGCAPELGGPSVFLPPPKKDTLMSRPEVYCVVVDVSHEDPEVKFNNYVDGV 461  
 DB 213 SSTCAASSTTIL-----FTTTPSPAD-IFLSKSNLTLCLVSNLATYV-TLNISWASQSG 266  
 QY 462 EVHNKATPREQYNTYRVVSVLTVLHQDMNGEKYCKSNKALPAPIETKISKAQO 521  
 DB 267 EPLEFKIKIMESHPRGTSAGKAVASVCVEDMNNRKEFCYTHRLPSPQKKFISK----- 322  
 QY 522 PRE-----POVYTLPPSHDEL-T-KNQVSLTCLVKGFPDIAVWESNGQ--PENNYKTT 573  
 DB 323 PNEVAKHPPAVYLLPPANEQNLRESATVTCGLVGFSPADISVQLQSGQLLPQKRYTS 382  
 QY 574 PVLVD--SDGSFPLYSKLTVDKSRWQGNVSCSMHVALNNHTYQKLSLSPGLQDET 631  
 DB 383 APMPGPAPGFYFTSHILTVEEWNNGSETYTCVGHALPHLVTERVDS-----TEG 437  
 QY 632 CAEADGDELGLMTT 646  
 DB 438 EYNAEEBGFENLWTT 452  
 RESULT 35  
 EPC\_MOUSE STANDARD; PRT; 421 AA.  
 AC P06336; P01856;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig epsilon chain C region.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84236092; PubMed=6329728;  
 RA Ishida N., Ueda S., Hayashida H., Miyata T., Honjo T.;  
 RT "The nucleotide sequence of the mouse immunoglobulin epsilon gene:  
 RT comparison with the human epsilon gene sequence.";  
 RN EMBO J. 1:1117-1123 (1982).  
 RN [2]  
 RP REVISIONS.  
 RA Honjo T.;  
 RL Submitted (Apr-1986) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 34-421 FROM N.A.  
 RX MEDLINE=83117774; PubMed=6818553;



RA Liu F.-T., Albrandt K., Sutcliffe J.G., Katz D.H.;  
 RT "Cloning and nucleotide sequence of mouse immunoglobulin epsilon  
 chain cDNA."  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:7852-7856(1982).  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X01857; CAA25977.1; -;  
 DR EMBL; X01857; CAA25978.1; -;  
 DR PIR; A02144; EHMS.  
 DR PIR; A02145; EHMS.  
 DR HSSD; P01854; 11GE.  
 DR InterPro; IPR007110; Ig-1-like.  
 DR InterPro; IPR003597; Ig\_c1.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00047; Ig; 4.  
 DR SMART; SM00407; IGc1; 2.  
 DR PROSITE; PS00835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.  
 FT NON TER 1 1  
 FT DOMAIN 1 90 CH1.  
 FT 198 197 CH2.  
 FT DOMAIN 198 304 CH3.  
 FT 305 421 CH4.  
 FT DISULFID 23 75 BY SIMILARITY.  
 FT DISULFID 121 180 BY SIMILARITY.  
 FT DISULFID 226 285 BY SIMILARITY.  
 FT DISULFID 330 392 BY SIMILARITY.  
 FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 421 AA; 47320 MW; 8F909E1F30A06B47 CRC64;  
 -----  
 Query Match 10.7%; Score 364.5; DB 1; Length 421;  
 Best Local Similarity 28.7%; Pred. No. 2,9e-16;  
 Matches 123; Conservative 75; Mismatches 163; Indels 67; Gaps 21;  
 -----  
 QY 223 FPLAFTVETKLNGSELMMQAEASASSKSMITFDLKNKE--VSVKRVTDQPKLQMGKLLPL 280  
 DB 29 FPNPTVTV-----WYSD--SLNMSTVNFPAISLKTTSQGVTSWGR--SAKNPTC 75  
 QY 281 HLTLPQALPYAGSGNLTL---ALEAKTGKIHQEVNLVWKATQLOKNTLCEVWGPTSP 336  
 DB 76 HVTNH---PSFNESRTILVRPNVITEPTLELHSSCDNNAHST---IQLYCFIYGHILN 129  
 QY 337 KLMISLKLKENEAKYKREKRPVWLVNPEAGMQLCLSSGOV--LLESIXKVLPTMSTFVE 395  
 DB 130 DVASVSMWMDREIDTTLAQT---VLIKKEGK--LASTCSKLNTTEQOMSESTFTCVT 183  
 QY 396 PKSCDKTTCPPCPAPPELLGSPVLPFPKPKDTLMIRTPREVTCVVVDV--SHEDPEVKF 454  
 DB 184 SQGVYDLAHTRCRPHN-PRGVITYLILPPSPD-LYQNGARKLTCLVLDSEKKNVNTW 241  
 QY 455 N-----WYVDGVEVNAKTKPREEOYNSTYRVSVLYTLVHODLMNGEKYCKKVS 503  
 DB 242 NOEKKTSSASQWY---TKHNN-----NATTSITSLIPVAKWIMGYGQICVD 288  
 QY 504 NKALPAPLEKTIISAKKQGPREFQVYTLPPSRDELTKNOVSLTCLVKGYPYSDIAVEMESN 563  
 DB 289 HPDFPKPIVRSITKTTPGORSAPBEVVFPPEEB--SEDKRTLTCLIONFFPEDISVOMLGD 347

QY 564 GQEPNN--YKTTPEVLDSGS---FPLYSKLTVDKSRMOQGNVSCSWMEALHN-HYTD 617  
 DB 348 GKLISSQHSHTTP-LKSNSSNGOFFFSRLLEVAKTLMTQRKQTCOVHIALQKPKLE 406  
 QY 618 KSLSLSPG 625  
 DB 407 KTIISTLDG 414  
 -----  
 RESULT 36  
 MOC\_MOUSE STANDARD; PRT; 455 AA.  
 ID MOC\_MOUSE  
 AC P01872;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig mu chain C region secreted form.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=81076590; PubMed=6255422;  
 RA Kawakami T., Takahashi N., Honjo T.;  
 RT "Complete nucleotide sequence of mouse immunoglobulin mu gene and  
 RT comparison with other immunoglobulin heavy chain genes."  
 RL Nucleic Acids Res. 8:3933-3945(1980).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=82051295; PubMed=6795090;  
 RA Goldberg G.I., Vanin E.F., Zolke A.M., Blattner F.R.;  
 RT "Sequence of the gene for the constant region of the mu chain of  
 RT Balb/c mouse immunoglobulin."  
 RL Gene 15:33-42(1981).  
 RN [3]  
 RP SEQUENCE FROM N.A. (MYELOMA TEPCL183).  
 RX MEDLINE=81165562; PubMed=6260591;  
 RA Aufreay C., Rougeon F.;  
 RT "Nucleotide sequence of a cloned cDNA corresponding to secreted mu  
 RT chain of mouse immunoglobulin."  
 RL Gene 12:77-86(1980).  
 RN [4]  
 RP SEQUENCE (MYELOMA PROTEIN MOPC 104E).  
 RX MEDLINE=79223904; PubMed=111247;  
 RA Kehry M.R., Sibley C.H., Fuhrman J.S., Schilling J.W., Hood L.B.;  
 RT "Amino acid sequence of a mouse immunoglobulin mu chain."  
 RL Proc. Natl. Acad. Sci. U.S.A. 76:2932-2936(1979).  
 RN [5]  
 RP REVISION (MOPC 104E).  
 RX MEDLINE=83075344; PubMed=6816276;  
 RA Kehry M.R., Fuhrman J.S., Schilling J.W., Rogers J., Sibley C.H.,  
 RT Hood L.B.;  
 RT "Complete amino acid sequence of a mouse mu chain: homology among  
 RT heavy chain constant region domains."  
 RL Biochemistry 21:5415-5424(1982).  
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoform=2;  
 CC Comment=During differentiation, B lymphocytes switch from  
 CC expression of isoform Membrane-bound to isoform Secreted;  
 CC Name=Secreted;  
 CC IsoId=P01872-1; Sequence=Displayed;  
 CC Name=Membrane-bound;  
 CC IsoId=P01873-1; Sequence=External;  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>



```

CC or send an email to license@slab-sib.ch).
CC -----
CC EMBL; V00827; ; NOT_ANNOTATED_CDS.
CC PIR; A02166; MHMS.
CC HSP; P01857; 1FC1.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003597; Ig_c1.
CC InterPro; IPR003006; Ig_MHC.
CC Pfam; PF00047; Ig; 4.
CC SMART; SM00407; IgC1; 2.
CC PROSITE; PSS0835; IG_LIKE; 4.
CC PROSITE; PSS0230; IG_MHC; 3.
CC Immunoglobulin domain; Glycoprotein;
CC Alternative splicing.
KW
FT NON_TER 1 1
FT DOMAIN 1 105 CH1.
FT DOMAIN 106 217 CH2.
FT DOMAIN 218 324 CH3.
FT DOMAIN 325 436 CH4.
FT DOMAIN 437 455 C-TERMINAL SEGMENT OF SECRETED FORM.
FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 28 89 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 136 199 BY SIMILARITY.
FT DISULFID 216 216 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 246 305 BY SIMILARITY.
FT DISULFID 293 293 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 353 415 BY SIMILARITY.
FT CARBOHYD 454 454 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .).
FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .).
FT CARBOHYD 442 442 S -> N (IN MOPC 104E).
FT VARIANT 78 78 S -> N (IN MOPC 104E).
FT VARIANT 101 101 H -> Q (IN MOPC 104E).
FT VARIANT 226 226 T -> N (IN TERC183 AND MOPC 104E).
FT VARIANT 258 258 N -> T (IN MOPC 104E).
FT VARIANT 258 258 N -> S (IN TERC183).
FT VARIANT 368 368 L -> K (IN TERC183 AND MOPC 104E).
SQ SEQUENCE 455 AA; 50101 MW; 4CB557CB602F9851 CRC64;

Query Match 10.7%; Score 364; DB 1; Length 455;
Best Local Similarity 22.8%; Pred. No. 3.5e-16;
Matches 122; Conservative 80; Mismatches 180; Indels 152; Gaps 17;

QY 144 ESPPGSSPVQCSRP-RGKRIQGGKTLVSQLELDSSGWTCTVLONOKKVEKIDIVVL 202
DB 3 QSPNVFPLVSCSPSLSKVLAMGCLARDPLPTTISFTWN----- 43
QY 203 AFQKASSIVYKKEGQVFEFSFLAFTVEKLTGSGELMQAERASSKSWITFDLKNKEVS 262
DB 44 -----YQNTTEVIO-----GRTPTLTGGAKYLA 68
QY 263 VVRVTQDPKLQW-----GKKLPLHLTLPLALPOVAGSGNLTALAEATGKL 308
DB 69 TQGVLLSPKSLIEGSDVYLCKIHYGGKRDHLVPIR-----AVAMM 110
QY 309 HOEVNLVY-----MATQLOKULTEVWGPPSPKMLSLKLENKAKYKAEKPYWVLN 362
DB 111 NPVNVFVPPRDGFGSPAPRPKSLICEATNFTPKPTVSMVLKDGKLVESGFTTDPVTJEN 170
QY 363 -----PEA-----GMMOCLLSDSGQVLLSNIKYVLPMTSTVEPRPSCDK 401
DB 171 KSTPTTYAVISTLTSEIDMLANLYTRVRDHRLGTLFK-----NV 212
QY 402 THTCPEPAPELLGGSVFLFPKPKDITMISTPEVTCVVVDVSHEDPEVFNMYVDGV 461
DB 213 SSTCAASPSTDL-----FTIPSPAD-IFLSKANLTLGVSNLAIVE-TLNISSMASGSG 266
QY 462 EVHNATKREGEQVNTYRVSVLYTLADMLNGKRYKKKVSINKALPAIEKTISCAKQ 521
DB 267 EPLETKIKIMESHPIGNTGSAKAVSVCVDEDMNRKEFVCTVTRDLSPQKFKISK--- 322

```

```

QY 522 PRE-----POVYTLPPSRDELTKNOVSLTCLVKGFPSYSDIAVESNSGQ--PENNYKTT 573
DB 323 PNEVHKHPAVYLLLPAPRGQLRESATVTLGLVFKFSPADISVQLQSGQLLPQEKYTS 382
QY 574 PVLV--SDGSFPLYSKLTVDKSRNQGVSFCSVMHEALHNHYTKSLISLSPG 625
DB 383 APMPEKGAPGFYFTHSILVTEENWNSGETYTCVVGHEALPHLVTERVTDKSTG 436

RESULT 37
MUC_HUMAN
ID MUC_HUMAN STANDARD; PRT; 454 AA.
AC P01871;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 19 mu chain C region.
GN IGHM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE OF 1-434 FROM N.A.
RX MEDLINE=90332450; PubMed=2115996;
RA Friedlander R.M., Nussenzweig M.C., Leder P.;
RT "Complete nucleotide sequence of the membrane form of the human Igm
RT heavy chain."
RL Nucleic Acids Res. 18:4278-4278 (1990).
RN [2]
RP SEQUENCE (WALDENSTROM'S MACROGLOBULIN GAL).
RX MEDLINE=75059123; PubMed=4803843;
RA Watanabe S., Barnikol H.U., Horn J., Bettram J., Hilschmann N.;
RT "The primary structure of a monoclonal Igm-immunoglobulin
RT (macroglobulin Gal.), II: the amino acid sequence of the H-chain (mu-
RT type), subgroup H III. Architecture of the complete Igm-molecule."
RL Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509 (1973).
RN [3]
RP REVISIONS (GAL).
RX MEDLINE=81066716; PubMed=6777162;
RA Mihaesco E., Barnikol-Watanabe S., Barnikol H.U., Mihaesco C.,
RA Hilschmann N.;
RT "The primary structure of the constant part of mu-chain-disease
RT protein B0F."
RL Eur. J. Biochem. 111:275-286 (1980).
RN [4]
RP SEQUENCE (WALDENSTROM'S OU), DISULFIDE BONDS, AND CARBOHYDRATES.
RX MEDLINE=74005511; PubMed=4742735;
RA Putnam F.W., Florent G., Paul C., Shinoda T., Shimizu A.;
RT "Complete amino acid sequence of the mu heavy chain of a human Igm
RT immunoglobulin."
RL Science 182:287-291 (1973).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=82059479; PubMed=6795593;
RA Rabbits T.H., Forster A., Milstien C.P.;
RT "Human immunoglobulin heavy chain genes: evolutionary comparisons of
RT C mu, C delta and C gamma genes and associated switch sequences."
RL Nucleic Acids Res. 9:4509-4524 (1981).
RN [6]
RP SEQUENCE OF 299-387 AND 438-454 FROM N.A.
RX MEDLINE=81077306; PubMed=6777778;
RA Dolby T.W., Devuono J., Croce C.M.;
RT "Cloning and partial nucleotide sequence of human immunoglobulin mu
RT chain cDNA from B cells and mouse-human hybridomas."
RL Proc. Natl. Acad. Sci. U.S.A. 77:6027-6031 (1980).
CC -I- MISCELLANEOUS: All 4 combinations of the S/G and V/G polymorphisms
CC at positions 192 and 216 have been observed in human mu chains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL: X17115; CAA34971.1; ALT\_SEQ.  
 DR EMBL: X57086; -, NOT\_ANNOTATED\_CDS.  
 DR HSSP: P01857; 1FCL.  
 DR Genew; HGNC:5541; IGHM.  
 DR MIM: 147020; -.  
 DR GlycoSiteDB: P01871; -.  
 DR GO: GO:0005624; C:membrane fraction; NAS.  
 DR GO: GO:0003823; F:antigen binding; NAS.  
 DR GO: GO:0006955; P:immune response; NAS.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003597; IG cl.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR Pfam: PF00047; Ig; 4.  
 DR SMART: SM00407; IGC1; 3.  
 DR PROSITE: PS50835; IG\_LIKE; 4.  
 DR PROSITE: PS00290; IG\_MHC; 3.  
 KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;  
 KW Polymorphism.

FT NON\_TER 1 1  
 FT DOMAIN 1 105 CH1.  
 FT DOMAIN 106 218 CH2.  
 FT DOMAIN 219 324 CH3.  
 FT DOMAIN 325 454 CH4.  
 FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN).  
 FT DISULFID 28 88  
 FT DISULFID 135 198  
 FT DISULFID 215 215 INTERCHAIN (WITH HEAVY CHAIN).  
 FT DISULFID 245 304  
 FT DISULFID 292 292 INTERCHAIN (WITH HEAVY CHAIN IN ANOTHER  
 SUBUNIT).  
 FT DISULFID 352 414  
 FT DISULFID 453 453 INTERCHAIN (WITH HEAVY CHAIN).  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD /FTID=CAR\_000219.  
 FT VARIANT 192 192 S->G.  
 FT VARIANT /FTID=VAR\_003903.  
 FT VARIANT 216 216 V->G (in dbSNP:12365).  
 FT VARIANT /FTID=VAR\_003904.  
 FT VARIANT /FTID=VAR\_003904.  
 SQ SEQUENCE 454 AA, 49556 MM, 21EC72BADC56522E CRC64;

Query Match 10.5%; Score 359.5; DB 1; Length 454;  
 Best Local Similarity 27.6%; Pred. No. 6.8e-16;  
 Matches 83; Conservative 62; Mismatches 111; Indels 45; Gaps 10;

QY 332 GPTSKMLSLKLNKKEAKVSKREKPVWVNLNPEAGMOCCLSDSCQVLE-SNIVLPTW 390  
 DB 173 GPTTKVSTLTIKESD-----WL---SQSMFCRDVHRLTLFOQNASNCVBDQ 219  
 QY 391 SRPVPRKSDCKHTCRPCAPARELLGAPSVFLFRPKDTLMSRTPEVTVCVVDVSHDP 450  
 DB 220 DTAIR-----VFALRPSP-FASIFLTSTKTLTCLVTLDTTDD- 254  
 QY 451 EVKFNWYDGVVHNAKTKPREEOYNSTYRVVSVLTVLHODLANKKEYCKVSNKALPAP 510  
 DB 255 SVTISWTHQNGEAVVHTHTNISSEHNATPSAVGASICEDDWNSGERFCTVHTDLPSP 314  
 QY 511 IEKTSKAKGP-REPOVYTLTPSRDELTKQVSLTCLVYKPFYSDAVESHSGGP-- 566  
 DB 315 LKQTSIRPGVALHAPDVYLLPAPAEQLNLRESATITCLVTGFSPADVFWOMRGOPLS 374  
 QY 567 ENNYTTPPVLD--SDGSFFLYSKLTVDRKRWQGNVFSVMHEALHNHYOKSLSP 624  
 DB 375 PEKYTSAWPEPQAPGRYFAHSILTVSEEWNTGETTCVVAHAELPRVTERIVDKST 434  
 QY 625 G 625

DB 435 G 435

RESULT 38  
 EPC\_HUMAN  
 ID EPC\_HUMAN STANDARD; PRT; 428 AA.  
 AC P01854;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig epsilon chain C region.  
 GN IGH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8316897; PubMed=6300763;  
 RA Seno M., Kurokawa T., Ono Y., Onda H., Sasada R., Igarashi K.,  
 RA Kikuchi M., Sugino Y., Nishida Y., Honjo T.;  
 RT "Molecular cloning and nucleotide sequencing of human immunoglobulin  
 epsilon chain cDNA.";  
 RL Nucleic Acids Res. 11:719-726(1983).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND VARIANT LEU-359.  
 RX MEDLINE=83001945; PubMed=6288268;  
 RA Max E.E., Battey J., Ney R., Kirsch I.R., Leder P.;  
 RT "Duplication and deletion in the human immunoglobulin epsilon genes.";  
 RL Cell 29:691-699(1982).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84236029; PubMed=6234164;  
 RA Flanagan J.G., Rabbits T.H.;  
 RT "The sequence of a human immunoglobulin epsilon heavy chain constant  
 region gene, and evidence for three non-allelic genes.";  
 RL EMBO J. 1:655-660(1982).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84207910; PubMed=6327276;  
 RA Ueda S., Nakai S., Nishida Y., Hisajima H., Honjo T.;  
 RT "Long terminal repeat-like elements flank a human immunoglobulin  
 epsilon pseudogene that lacks introns.";  
 RL EMBO J. 1:1539-1544(1982).  
 RN [5]  
 RP PRELIMINARY SEQUENCE (MYELOMA PROTEIN ND).  
 RA Benrich H.H., Johansson S.G.O., von Bahr-Lindstrom H.;  
 RL (in) Bach M.K. (eds.);  
 RT Immediate hypersensitivity: modern concepts and developments, pp.1-36,  
 RL Marcel Dekker, New York (1978).  
 RN [6]  
 RP SEQUENCE OF 1-40; 68-114 AND 427-428 FROM N.A.  
 RX MEDLINE=83065234; PubMed=6815656;  
 RA Kenten J.H., Mølgard H.V., Houghton M., Derbyshire R.B., Viney J.,  
 RA Bell L.O., Gould H.J.;  
 RT "Cloning and sequence determination of the gene for the human  
 immunoglobulin epsilon chain expressed in a myeloma cell line.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:6661-6665(1982).  
 RN [7]  
 RP 3D-STRUCTURE MODELING.  
 RX MEDLINE=87089848; PubMed=3796618;  
 RA Padlan E.A., Davies D.R.;  
 RT "A model of the Fc of immunoglobulin E.";  
 RL Mol. Immunol. 23:1063-1075(1986).  
 CC -I- SIMILARITY: Contains 4 immunoglobulin-like domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>



```
CC or send an email to l1cense@sb-ib.ch).
CC EMBL; L00022; AAB59424.1; ALT_INIT.
DR PIR; A23771; EHHU.
DR PDB; 1ICE; 15-JUL-92.
DR PDB; 1FP5; 30-JAN-02.
DR PDB; 1G84; 16-MAY-01.
DR PDB; 1OOV; 18-SEP-02.
DR Genew; HGNC:5522; IGHE.
DR MIM; 147180.
DR GO; 0003823; F:antigen binding; NAS.
DR GO; 0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00407; IGc1; 4.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW 3D-structure; Repeat.
FT NON_TER 1 1
FT DOMAIN 6 103 IG-LIKE 1.
FT DOMAIN 112 210 IG-LIKE 2.
FT DOMAIN 214 318 IG-LIKE 3.
FT DOMAIN 324 423 IG-LIKE 4.
FT DISULFID 14 14 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 15 105 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 29 85 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 121 121 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 135 193 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 209 209 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 239 299 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 345 405 INTERCHAIN (WITH A HEAVY CHAIN).
FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .).
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .).
FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .).
FT CARBOHYD 359 359 W -> L (POSSIBLE POLYMORPHISM).
FT VARIANT 359 359 /FTid=VAR_003885.
FT STRAND 110 110
FT STRAND 113 117
FT HELIX 122 124
FT TURN 125 125
FT STRAND 130 141
FT STRAND 146 150
FT STRAND 155 155
FT STRAND 159 161
FT STRAND 165 166
FT TURN 168 169
FT STRAND 172 181
FT HELIX 182 186
FT TURN 187 188
FT STRAND 192 196
FT TURN 198 199
FT TURN 201 202
FT STRAND 205 207
FT STRAND 218 222
FT TURN 226 229
FT TURN 231 232
FT STRAND 236 242
FT STRAND 252 252
FT STRAND 255 257
FT TURN 258 259
FT STRAND 260 260
FT TURN 265 266
FT STRAND 271 274
FT TURN 275 276
FT STRAND 277 285
FT HELIX 288 293
FT TURN 294 294
```

```
FT STRAND 297 302
FT TURN 304 305
FT STRAND 310 314
FT STRAND 322 322
FT STRAND 325 329
FT HELIX 334 337
FT STRAND 340 348
FT STRAND 350 351
FT STRAND 356 360
FT TURN 366 369
FT STRAND 374 374
FT STRAND 377 378
FT TURN 380 381
FT STRAND 384 385
FT STRAND 387 393
FT HELIX 394 398
FT TURN 399 400
FT STRAND 404 408
FT STRAND 418 421
SQ SEQUENCE 428 AA; 47019 MW; 25C4CA072AA558A0 CRC64;

Query Match
Best Local Similarity 10.5%; Score 359; DB 1; Length 428;
Matches 119; Conservative 69; Mismatches 155; Indels 98; Gaps 18;

QY 267 TDDPKL-----QWCKLPLHLT-----LPOALPOVAGSGNL---TLALEAKTGKL 308
DB 3 TQSPSVFPLTRCKKNIPSNATSVTLGCLATGYFPEPVAVMTDTSLSNGTMTLPAITLTL 62
QY 309 --HGEVNLVVRATQLOKNLTCEV-WGPTSPKMLSLKENKAKYSKRE---KPVWVL 361
DB 63 SGHYATISLTVSGAMAKQMFQVRVAHTPSSTDW----VDNKTPSVCSRDPTPTVKIL 117
QY 362 N-----PEAGMWQCLLS-----DSGQVLLSNNIKVLEPTWSTPVVPRKC- 399
DB 118 QSSCDGGHFPETIQLLCLVSGYTPGTINITWLEBDQVMDVD---LSTASTQGEELAS 173
QY 400 -----DKTHTC-----PPCPAPELLGSPVFLPPPKPDITMI 432
DB 174 TQSELTLSQKMWLSRITTCQVTVGHTFEDSTKKC-ADSNPRGVSAVLSRSPFD-LFI 231
QY 433 SRTPVTCVVVDVSHEDPEVKFNWYVDGVEVNAKTKPRREQYNSTYRVSVLTVLDHDM 492
DB 232 RKSPTITGLVVDLAVSKGTVALTWGRASGKPVNSTRKEBKQRNGLTVTSTLPVGTBDW 291
QY 493 LNKETKKKVSNAKLPAPLEKTISSAKQPREPOVYTL-----PPSRDELTKNQVSLTCL 547
DB 292 IEGETYYQCRVTHPHLPALMRSTTKTSGPRAPEVYAFATPEWPGSRDKRT-----LACL 346
QY 548 VKGFPYSDIAVEWSEN--GQPENNYKTTTPVLDSGSPFLYSKLVLDKSRMQQGVFSGS 605
DB 347 IQNFWPEDISVQWMLNEVQLPDPARSTTQPRKTKSGGFVFSRLVTAWEQKDEFICR 406
QY 606 VMHEALHNHYT-QKSLSLSPG 625
DB 407 AVHEAASPQTYQRAVSYNPG 427

RESULT 39
MUCM RABIT STANDARD; PRT; 479 AA.
AC P04231.
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig mu chain C region membrane-bound form.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A. (A2 ALLOTYPED).
RX MEDLINE=8408930; PubMed=6418803;
```



```

RA Bernstein K.E., Alexander C.B., Reddy E.P., Mage R.G.
RT "Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain
RL J. Immunol. 132:490-495(1984).
CC -I- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment-During differentiation, B lymphocytes switch from
CC Name=Membrane-bound;
CC IsoId=P04221-1; Sequence=Displayed;
CC Name=Secreted;
CC IsoId=P03988-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K01357; AAA1293.1; -
DR PIR: A02165; MHRM.
DR HSP: P01842; 7FAB.
DR InterPro: IPR007110; Ig_1like.
DR InterPro: IPR003597; Ig_c1.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_4.
DR SMART: SM00407; IgC1.2.
DR PROSITE: PS00835; Ig_LIKE; 4.
DR PROSITE: PS00290; Ig_MHC; 3.
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing; Transmembrane.
FT NON_TER 1 1
FT DOMAIN 1 106 CH1.
FT DOMAIN 107 222 CH2.
FT DOMAIN 223 327 CH3.
FT TRANSMEM 328 458 CH4.
FT DISULFID 459 476 POTENTIAL.
FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 28 90 BY SIMILARITY.
FT DISULFID 137 200 BY SIMILARITY.
FT DISULFID 219 219 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 249 308 BY SIMILARITY.
FT DISULFID 296 308 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 356 418 BY SIMILARITY.
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 479 AA, 52351 MW, 689C637A47BE19FC CRC64;

Query Match 10.5%; Score 358.5; DB 1; Length 479;
Best Local Similarity 24.6%; Pred. No. 8.4e-16;
Matches 121; Conservative 83; Mismatches 175; Indels 113; Gaps 21;

QY 223 PPL-----AFYVEKLTSGSELMWQERASSSKSMITPPLKXKEVSVKAVTDDPKLQMGKK 277
DB 9 YPLVSCGALTDGNLVAMGCLARDPLPSVVFPSW-SFK-NNSEISSRTVTFPPVKKRQDK 66
QY 278 -----LPLHLTLPOAL-----PGYA 292
DB 67 YMAISQVLVPKDVLYQGTBEYIVCVQHSNSNRDLRVFPVDSLEPPNVSVPFIPRDSFS 126
QY 293 GSGNLTLLALEAK-TGKLHQEVNLVVMRATQ-----LQKNLTCEVWGPTSPKMLSLKLE 345
DB 127 GSGTKSKSLIKQATGFPSPKQISVSWLRDQKVESGVLRKPYEATKKGAPRTFSLSMLT 186
QY 346 NKEANVSREKPKVWVLANPAGMOCGLDSDGVLLIESIKVLPFWSTVPEPKSCDKHTTC 405
DB 187 ITESD-----WL-----SQSLYTCRVDRHG-IPFDKNVSM-----SSSEC 219

```

```

QY 406 PRCAPELLGSPVFLPPPKDTLMISRTPEVTCVVVDVSHEDPEVKFMYVDGEVHN 465
DB 220 STTSP-----GIQFPIAPSPADI-FLSKGARLCLVTDLTYY-SLNIW-----ASHN 268
QY 466 AKT-----KPREQYNSTYRVSVLVVLDHMDLNGKEYKKCKVSKNALPAIEKTSKAG 520
DB 269 GKALDTMNITESHSPNTPFAMGEASVCADMEGSEQFTCTVTHADLPFLKHTISKSRE 328
QY 521 QPRP-POVYTLPPSRDLT-KNOVSLTCLYKGFYPSDIAVWESNGQP--ENNYKTPPV 576
DB 329 VAKPPAVVYLPAPREQLVRESATVTCVLVGFSPADVPVQMOQRGQPLSSDKVTAPA 388
QY 577 LD--SDSFFLYSKLVYDKSRMOQGVFSCSVMEALHNHYTKSLSPGLQDETCAE 634
DB 389 PERAPGLYTHSLTYTTEEDMNSGERTFCVGHFALPHWTERTVKS-----TEGEVG 443
QY 635 AQDGLDGLMTT 646
DB 444 AEEGPFENLWTT 455

RESULT 40
EPC_RAT ID EPC_RAT STANDARD; PRT; 429 AA.
AC P01855;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig epsilon chain C region.
OC Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A. (IMMUNOCYTOMA IR2).
RC STRAIN=LOU/C/MSL.
RX MEDLINE=83064537; PubMed=6292865;
RA Hellman L., Petersson U., Engstrom A., Karlsson T., Bennich H.;
RT "Structure and evolution of the heavy chain from rat immunoglobulin
RT E.";
RL Nucleic Acids Res. 10:6041-6049(1982).
RN [2]
RP SEQUENCE OF 168-342 FROM N.A. (MYELOMA IR162).
RX MEDLINE=83182019; PubMed=6820340;
RA Kindsvogel W.R., Reddy E.P., Moore J.M., Fauser C.H. Jr.;
RT "A cloned cDNA probe for rat immunoglobulin epsilon heavy chain:
RT construction, identification, and DNA sequence.";
RL DNA 1:335-343(1982).
RN [3]
RP SEQUENCE OF 205-306 FROM N.A.
RX MEDLINE=82174576; PubMed=680238;
RA Hellman L., Petersson U., Bennich H.;
RT "Characterization and molecular cloning of the mRNA for the heavy
RT (epsilon) chain of rat immunoglobulin E.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1264-1268(1982).
CC -I- SIMILARITY: Contains 4 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J00744; AAA41379.1; ALT_INIT.
DR PIR: A93442; EHRT.
DR HSP: P01854; 11GE.
DR InterPro: IPR007110; Ig_1like.
DR InterPro: IPR003597; Ig_c1.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_4.

```



```
DR SMART, SM00407, IGC1, 1.
DR PROSITE: PS00835; IG_LIKE, 4.
DR PROSITE: PS00290; IG_MHC, 3.
KW Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON_TER 1 1
FT DOMAIN 6 89 IG-LIKE 1.
FT DOMAIN 103 201 IG-LIKE 2.
FT DOMAIN 205 305 IG-LIKE 3.
FT DOMAIN 314 414 IG-LIKE 4.
FT CONFLICT 168 168 R -> N (IN REF. 2).
FT CONFLICT 308 308 P -> L (IN REF. 2).
SQ SEQUENCE 429 AA; 48671 MW; D2970B34EFA7B0 CRC64;

Query Match
Best Local Similarity 27.8%; Pred. No. 1,1e-15;
Matches 101; Conservative 69; Mismatches 114; Indels 79; Gaps 16;

QY 324 KNLTCEWGPSTPKMLSLKLENKAKVSKREKPVVVLNPEAGM----- 367
DB 74 KNLTCHV--THASTFVS-----DLTIRAPVITKPTVDLHSSCDPNAFHTIOL 123
QY 368 -----WQCLSD-----SGVYLESNIKVLEPTWS---TPVEPKSCDKT 402
DB 124 YCFVYGHIONDYSIHW--LMDDRKIYETHAQNVLKEBEKGLASTYSRLNITQQQWMSBST 181
QY 403 HICRP-----PCPAPRLGGPSVLEFPKPKXTLMSRTEVTCVVDVSHED 449
DB 182 FTCKKTSGEBNVAHTRRCSDDP--PRGVITTYLIPSPDL--LVENGPKLTLCLVDLSESE 239
QY 450 PEVKEFMYVGVGVENNAKTRPREEQNSTYRVVSVTLVHODMLNKEYCKVSNKALPA 509
DB 240 -NITVWVERKKKSISASORSKHNATTSITSLPVAKMIKEBGVQCRVDHHPFK 298
QY 510 PLEKTSKAKGQPREPOVYT--LPSPDELTKQVSLTCLVKGFYPSDIAVEW--ESNGQP 566
DB 299 FYRSITKAPGKSABEVYVFLPPEEBE--KDKRTLTCLIQHFFPEDISVQWLOQSKLIP 356
QY 567 ENNYKTPPVLDSDG---FELYSKLTVDKSRWQGNVSCSVMEHALIN--HYTQSLSL 622
DB 357 KQHSHTTTP-LKYNNGSNQRFIFSRLEVTALMTQKPTCRVHIEALREPKLERTISK 415
QY 623 SPG 625
DB 416 SLG 418

RESULT 41
MUC_MESAU
ID MUC_MESAU STANDARD; PRT; 454 AA.
AC P06317;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE 15 mu chain C region.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OC NCBI_TaxId=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85297761; PubMed=2994005;
RA McGuire K.L., Duncan W.R., Tucker P.W.;
RT "Phylogenetic conservation of immunoglobulin heavy chains: direct
RT comparison of hamster and mouse Cmu genes.";
RL Nucleic Acids Res. 13:5611-5628(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X02804, CAA26574.1; -.
DR HSSP, F01854, IIGS.
DR InterPro, IPR007110, Ig-1-like.
DR InterPro, IPR003597, Ig-cl.
DR InterPro, IPR003006, IG_MHC.
DR Pfam, PF00047, ig; 4.
DR SMART, SM00407, IGC1, 2.
DR PROSITE, PS00835; IG_LIKE, 4.
DR PROSITE, PS00290; IG_MHC, 3.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON_TER 1 1
FT DOMAIN 106 218 CH1.
FT DOMAIN 219 324 CH2.
FT DOMAIN 325 454 CH3.
FT DOMAIN 13 13 CH4.
FT DISULFID 27 88 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 135 198 BY SIMILARITY.
FT DISULFID 215 215 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 245 245 BY SIMILARITY.
FT DISULFID 292 292 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 352 414 BY SIMILARITY.
FT DISULFID 453 453 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT CARBOHYD 45 45 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 210 210 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 257 257 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 441 441 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 454 AA; 50196 MW; 4DA2134612BE1469 CRC64;

Query Match
Best Local Similarity 26.0%; Pred. No. 1,2e-15;
Matches 120; Conservative 82; Mismatches 167; Indels 93; Gaps 19;

QY 223 PFLA-----FTVEKLTSGGELMWMQERASSKSWITFDLKNKEVSKRVTDPKLQWGGK 277
DB 8 PFLVCSBPISDENLVAMGCLARDPLPSISFSW---NYQNSKSEVNGVGFPTLRMGSK 64
QY 278 -----LPLHLTPQALPYAGSGNLTLLAEATGTLHGEVNLVWRAATOLQCN----- 325
DB 65 YAATSQVFLP-----PKSVLE--GSDXY-LVCKVHAGNTNDRVPIVVTENMNVSVF 116
QY 326 -----LTCVWGPSTPKMLSLKLENKAKVSKREKPVVVLNPEAGMQ 369
DB 117 VSRDAPSGPAPRKSRRLFCESANFSPKQITVSMRDGKPVASGFTTEFV---TPB----- 168
QY 370 CLLSDSGQ-----VLLESNIKVLPTWSTPVEBKSC---DKTHTCPCPAPRL 414
DB 169 ----DRGSGPRTYKVIITLTITSDMLNLSYTCVDRHGLTFMWKGVSTCAASPTDIO 224
QY 415 GGPSTFLPPPKPXT--LMSRTEPVTGVVDVSHDEPVKRNWYDGVENNAKTRPBE 473
DB 225 A-----PPIPSFVGIFLNKSAITLCLVTMLATYD--TLNLSWSSRSSEPEETKTLRES 277
QY 474 QVNSTYRVVSVTLVHODMLNKEYCKVSNKALPAPIEKTISKAKGQPRE-----POVY 528
DB 278 HENGTFSAIGEANVCVEWDSKGEVCTVTHRDLPSPQKRTISK-----PRENNKTPPAVY 333
QY 529 TLPPSRDEL-TKNQVSLTCLVKGFYPSDIAVEWESNGQP--ENNYKTPPVLDSDG---F 583
DB 334 QQPLARQLILRESATVCLVKGFGPADIPQWLDRGQPLQDKVTSAPMREBPQAPHLV 393
QY 584 FYLSKLTVDKSRWQGNVPSGVMEALAHNYTQSLSLSG 625
DB 394 FTHSVTLVTEBWNSGETTYCVVGHFALPMWERTVDNRSTG 435

RESULT 42
```



```

MUCB_HUMAN          STANDARD;          PRT;          391 AA.
ID   MUCB_HUMAN      P04220;
AC   P04220;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ig mu heavy chain disease protein (BCR).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=84184186; PubMed=6425189;
RA   Barnikol-Watanabe S., Mihaesco E., Mihaesco C., Barnikol H.U.,
RA   Hilschmann N.;
RT   "The primary structure of mu-chain-disease protein BCR. Peculiar
RT   amino-acid sequence of the N-terminal 42 positions.";
RL   Hoppe-Seyler's Z. Physiol. Chem. 365:105-118(1984).
CC   -1- MISCELLANEOUS: This protein has no V region homology or CH1
CC   region.
CC   PIR: A02163; MHUBT.
CC   HSSP: P01857; 1PCL.
CC   DR   GO: GO:0005624; C:membrane fraction; NAS.
CC   DR   GO: GO:0003823; F:antigen binding; NAS.
CC   DR   GO: GO:0006955; P:immune response; NAS.
CC   DR   InterPro: IPR007110; IG_1like.
CC   DR   InterPro: IPR003597; IG_CL.
CC   DR   InterPro: IPR003006; IG_MHC.
CC   DR   Pfam: PF00047; Ig_3.
CC   DR   SMART: SM00407; IgC1; 2.
CC   DR   PROSITE: PS50835; IG_LIKE; 3.
CC   DR   PROSITE: PS00290; IG_MHC; 3.
CC   KW   Immunoglobulin domain; Immunoglobulin C region.
FT   DOMAIN          1      42      PRE-C-PART (NO V REGION HOMOLGY).
FT   DOMAIN          43      155     CH2.
FT   DOMAIN          156     261     CH3.
FT   DOMAIN          262     391     CH4.
FT   SEQUENCE        391 AA; 43057 MW; 9100843AF0CF021A CRC64;

Query Match          10.4%; Score 353.5; DB 1; Length 391;
Best Local Similarity 27.8%; Pred. No. 1.3e-15;
Matches 84; Conservative 61; Mismatches 110; Indels 47; Gaps 11;

QY  332 GPTSPKMLSLKLEKKAQVSKREKPVVNLNPEAGMOCLSDSGQVLLSENIKVL--PT 389
DB  110 GPTTYKVTSLTIKESD-----WL---GQSMFCRDYHNG-LTFQGNASMGCPD 155
QY  390 WSTPVEPKSCDKHTCCPCAPBELIGPSVFLFPKPKDTLMISRTPEVTCVVVDVSHED 449
DB  156 QDTAIR-----VFALPPS--FASIFLTSTCLTCLVDTLTYD 191
QY  450 PEVKFNWVVDGVNNAKTPREBOYNSTYRVSVLTVLHODMNGKVKYKKNKALPA 509
DB  192 -SVTISMRQGEAVKHTNISHPNATFSAVGEASICEPDWSGEFTCTVHTDIPS 250
QY  510 PIEKTIKAKQCP-REPQVYTLPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQCP- 566
DB  251 PLKOTISPKQVALHRPVVILLPRAREQDLNRESATITCLVTSPPAVFQVMQMGKQPL 310
QY  567 -ENNYKTPPVLD--SDGSFLYSKLTVDKSRMGOQGVFSCVMEALHNYTQKSLS 623
DB  311 SPEKVTSAFMPEDQAPGRVFAHSITLVSEEMWTGETYTCVVAHEALPNVTERTVDKS 370
QY  624 PG 625
DB  371 TG 372

```

```

DT  23-OCT-1986 (Rel. 02, Created)
DT  23-OCT-1986 (Rel. 02, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ig mu chain C region secreted form.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A. (A2 ALLOTYPE).
RX   MEDLINE=84088930; PubMed=6418803;
RA   Bernstein K.E., Alexander C.B., Reddy E.P., Mage R.G.;
RT   "Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain
RT   of Vha2 allotype: comparisons with Vha1 and membrane mu sequences.";
RL   J. Immunol. 132:490-495(1984).
CC   -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC   -1- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Comment=During differentiation, B lymphocytes switch from
CC   expression of isoform Membrane-bound to isoform Secreted;
CC   Name=Secreted;
CC   IsoId=P03988-1; Sequence=Displayed;
CC   Name=Membrane-bound;
CC   IsoId=P04221-1; Sequence=External;
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@sib-sib.ch).
CC   -----
DR   EMBL: K01357; -; NOT_ANNOTATED_CDS.
DR   PIR: A02164; MHRB.
DR   HSSP: P01842; 7PAB.
DR   InterPro: IPR007110; IG_1like.
DR   InterPro: IPR003597; IG_CL.
DR   InterPro: IPR003006; IG_MHC.
DR   Pfam: PF00047; Ig_4.
DR   SMART: SM00407; IgC1; 2.
DR   PROSITE: PS50835; IG_LIKE; 4.
DR   PROSITE: PS00290; IG_MHC; 3.
KW   Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW   Alternative splicing.
FT   NON_TER        1
FT   DOMAIN          1      106     CH1.
FT   DOMAIN          107     222     CH2.
FT   DOMAIN          223     327     CH3.
FT   DOMAIN          328     458     CH4.
FT   DISULFID        14      14      INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT   DISULFID        28      90      BY SIMILARITY.
FT   DISULFID        137     200     INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT   DISULFID        219     219     BY SIMILARITY.
FT   DISULFID        249     296     INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT   DISULFID        296     308     BY SIMILARITY.
FT   DISULFID        356     418     INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT   CARBOHYD        46      46      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        114     114     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        212     212     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        261     261     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        277     277     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        284     284     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        445     445     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SEQUENCE        458 AA; 49897 MW; F330D6A3B28E4864 CRC64;

Query Match          10.2%; Score 349; DB 1; Length 458;
Best Local Similarity 24.6%; Pred. No. 3.2e-15;
Matches 116; Conservative 80; Mismatches 167; Indels 108; Gaps 20;

QY  223 FPL-----AFVTEKLTSSGELTMQAEKASSSKMITFDLKNKQSVSVKRVYQDPKLGK 277
DB  9 YPLVSCGALTLDNVLVAMGCLARDPLSSSVTFSM-SFK-NNSETISHTVTRTFPVVKKRGDK 66

```



```

QY 278 -----LPLHLTPOAL-----POYA 292
DB 67 YWATSGVLPKDVLOGTEBYLVCKVQHSNSNDLAVSPVDSLPVSPVFPSPDPS 126
QY 293 GSGNLTALAEAK-TGHLHGVNLVNRATQ-----LQKNLTCEWGPSPKMLSLKLE 345
DB 127 GSGTRKSRLICQATGSPKQISVSMRLDQKVESGLTKPVEAKTGAGAPATFSSMLT 186
QY 346 NKEAKVSKREKPVVNLNPAEAMQCLSDSGVLLSNNIKLPTWSTPVPKSCDKTHRC 405
DB 187 ITESD-----WL-----SOSLYTCVDRHG-IFPKNVSM-----SSBC 219
QY 406 PCCPAPELLGSPSVFLFPKPKDITLMSRTPBEVTCVVVDVSHEDPEVKFNMYVDGVEVNN 465
DB 220 STTSPS-----GIQVFPIASPADT-FLSKSARLICLVLTLYTG-SLNSW-----ASN 268
QY 466 AKT-----KPREQVNSTYRVVSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTISKAKG 520
DB 269 GKALDTHNMITESHPRATFSAMGEASVCAEDMESGQFCTVTHADLPPLKHTTISKSR 328
QY 521 QPRE-POVYTLPPSRBELT-KNOVSLTCLYKGFYPSDIAVEMESNGQP--ENNYKTPPV 576
DB 329 VAKHPAVYVLPAPARQLVRESATVTCVYKGFSPADVFQWQORQQLSSDKYVTSAPA 388
QY 577 LD--SDGSFFLYSKLTVDKSRMQQNVFSCSVNHEALHNHYTKSLSPG 625
DB 389 PEPQAFGLYFTHSTLTVTBEDMNSGETFTCVGHEALPMTWERTYVDKSTG 439

```

## RESULT 44

MUC\_SUNMU STANDARD; PRT; 457 AA.

```

AC P20768;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE IG mu chain C region.
OS Suncus murinus (House shrew) (Musk shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Insectivora; Soricidae; Crocidurinae; Suncus.
OX NCBI_TaxId=9378;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89232144; PubMed=2497033;
RA Ishiguro H., Ichihara Y., Namikawa T., Nagatsu T., Kurosawa Y.;
RT "Nucleotide sequence of Suncus murinus immunoglobulin mu gene and
RT comparison with mouse and human mu genes.";
RL FEBS Lett. 247:317-322 (1989).

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X13920; CAA32113.1; ALT_INIT.
CC PIR, S03961; S03961.
CC HSSP, P01842; 7PAR.
CC InterPro, IPR007110; Ig-1-like.
CC InterPro, IPR003597; Ig_c1.
CC InterPro, IPR003006; Ig_MHC.
CC Pfam, PF00047; Ig_4.
CC SMART, SM00407; Igc1; 3.
CC PROSITE, PS00835; Ig_Like; 4.
CC PROSITE, PS00290; IG_MHC; 3.
CC Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON_TER 1 105 CH1.
FT DOMAIN 106 220 CH2.

```

```

FT DOMAIN 221 326 CH3.
FT DOMAIN 327 457 CH4.
FT DISULFID 13 13 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 27 89 BY SIMILARITY.
FT DISULFID 136 200 BY SIMILARITY.
FT DISULFID 217 217 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 248 307 BY SIMILARITY.
FT DISULFID 295 295 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 355 417 BY SIMILARITY.
FT DISULFID 456 456 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 457 AA; 50074 MW; 56C8C086DA4462E9 CRC64;

```

Query Match 10.2%; Score 348; DB 1; Length 457;

Best Local Similarity 23.7%; Pred. No. 3,7e-15; Matches 109; Conservative 90; Mismatches 160; Indels 100; Gaps 18;

```

QY 204 FQKASSIYKKEGEVESEFPPLAFTVEKLTGSGELMWQAE-ASSSKSWITFDLKN----- 258
DB 43 FKNSSSI-----SQNIYNPEVFTGKYMATSGVLLSTAILQSTDDYITCHTHTTGE 97
QY 259 KEVSXKRTQDPKLOMGKKLPLHLTL--POLPOYAGSGNLTALAEATGKLHGVNLV 316
DB 98 KEKVV-----ELQVTELPNVAIFVP--PRNSFSGN-----HPTSQLI 135
QY 317 MKATQLOKLTCEV-----GPTPKMLSLKLENKAKVS 352
DB 136 COASGFSPTRYTMSLQGEPRQPSLVSTSAVEAPKSGPTTFVIRLITTE----- 191
QY 353 KKEKPVWNLNPAEAMQCLSDSGVLLSNNIKVPTWSTPVEPKSCDKHTCPCPAPE 412
DB 192 -----WLSQRE--FTCOALHKG-LTFQKNVSV-----CWGDD 221
QY 413 LLGGSVYVLPFPKPKDITLMSRTPBEVTCVVVDVSHEDPEVKFNMYVDGVEVNNATKPRE 472
DB 222 TSTGTVFLPPTPAN-LFTQSQALTLVGLATYD-SLISWRONGEALQTHVNISE 279
QY 473 EGVNSTYRVVSVLTVLHQMVLNGEKYCKVSNKALPAPIEKTISKAKQPRE-POVYTL 531
DB 280 SHPNSTFPAKGHASVCREMESGEKFTCTVQHSIDLPKSLSPKPVANDPPEVFLP 339
QY 532 PSRDEL-TKNQVSLTCLYKGFYPSDIAVEMESNGQP--ENNYKTPPVLD--SDGSFFLY 586
DB 340 PAQEQGLKLRRESASTICLVKDFSPDPVFQWQHGGPVPKHYVTSNPPREPQNPGLYFVH 399
QY 587 SKLTVDKSRMQQNVFSCSVNHEALHNHYTKSLSPG 625
DB 400 SILTVSEKDWMSGESFSCVVGHEALPLSVTERKAVDKTSG 438

```

## RESULT 45

MUC\_CANFA STANDARD; PRT; 450 AA.

```

AC P01874;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG mu chain C region.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxId=9615;
RN [1]
RP SEQUENCE OF 1-177 (MOO).
RX MEDLINE=8007682; PubMed=117299;
RA McCumber L.J., Capra J.D.;
RT "The complete amino-acid sequence of a canine mu chain.";
RL Mol. Immunol. 16:565-570 (1979).

```



```

RN [2]
RP SEQUENCE OF 178-450 (MOO).
RX MEDLINE=78180587; Pubmed=653360;
RA Maaserman R.L., Capra J.D.;
RT "Antino acid sequence of the FC region of a canine immunoglobulin M:
RL Interpecies homology for the IGM class.";
Science 200;1159-1161 (1978).
DR PIR; A93131; MHDG.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KM Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER
SQ SEQUENCE 450 AA; 48895 MW; 9D46DDA9D1012F5D CRC64;

Query Match 10.0%; Score 340; DB 1; Length 450;
Best Local Similarity 25.9%; Pred. No. 1,2e-14;
Matches 100; Conservative 76; Mismatches 142; Indels 68; Gaps 15;

QY 278 LPLHLTLPLAL-----PQYASGN---LTALAEKTKLHGVNLVVRATOLQKULT-- 327
DB 102 LPVMTLTPPEVSGFIPRDAFPGBRKSQLICQASGFSPPQVWSL--RDGQIESGVTTN 159
QY 328 -----CEWGPRTSPRLMLSLLENKAKVSKREKVVWLNDEAGMQLSDSQVLLS 382
DB 160 EVZAZAKSGSPPTTYVVTSMULT-----QEBAML--SQSFETCKVSHRG-LTFQO 205
QY 383 NIKVLPWTSTPEPKSCDKHTHTCPAPPELLGSGSVLPFPKPKDTLMISRTPEVTCV 442
DB 206 NASSMCTSDQV-----GISITITPPS--PASIFNTKSAKLSTLV 243
QY 443 VDVSHEDEYKFMNVYDGEVHNNAKTKPREQYNSTYRVSVLVYHMDWLNKREYKCV 502
DB 244 TDLATYD-SVTISWIRENGALKTHNTINSESHPNCTFSAMGBATVCVEWESGEOFTCTV 302
QY 503 SNKALPAIEKTISSAKG-QPREPOVYTLPPSRDEL-TKNQVSLTCLVKGYSPIADIENV 560
DB 303 THTDLPVALKQITISPKVAAVAMPSPVYVLPSPREQLDRESATLSCLVTGYSPDPVFQW 362
QY 561 ESNQO--PENNYKTPPVLD--SDGSFLYSKLTVDKSRMOQGVFSGSVMEALHNHYT 616
DB 363 VOKGQPVPPDSVYSAPMPPEQAPSLYFAHSILTVSEEMNAGETTYCVVAHESLPRVY 422
QY 617 QKSLSLSPG-----LQDDETCAE 634
DB 423 ERSVDKSTGKPTLVYVSLVLSDTAGZ 448

RESULT 46
ALC_RABIT STANDARD; PRT; 299 AA.
ID 21-JUL-1986 (Rel. 01, Created)
AC P01879;
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DB Ig alpha chain C region (Fragment).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
NCBI_TaxID=9986;
RN SEQUENCE FROM N.A.
RP MEDLINE=84144059; Pubmed=6322114;
RA Knight K.L., Marens C.L., Stoklosa C.M., Schneiderman R.D.;
RT "Genes encoding alpha-heavy chains of rabbit IgA: characterization of
RT cDNA encoding IgA-g, subclasses alpha-chains.";
Nucleic Acids Res. 12:1657-1670 (1984).
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body

```

```

CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -1- MISCELLANEOUS: This immunoglobulin belongs to the IgA-G subclasses.
CC It was isolated from a rabbit homozygous FOR A2, N60, DE12, 15,
CC F71, G75 heavy chain haplotype.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC EMBL; X00353; CAA25100.1; -.
DR PIR; A02174; AHRB.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00407; IGc1; 2.
DR PROSITE; PS50835; IG_LIKE; 2.
DR PROSITE; PS00290; IG_MHC; 2.
KM Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON_TER
FT DOMAIN 1 167 IG-LIKE 1.
FT DOMAIN 174 276 IG-LIKE 2.
SQ SEQUENCE 299 AA; 32256 MW; 2512FE3F62B9A223 CRC64;

Query Match 8.9%; Score 305; DB 1; Length 299;
Best Local Similarity 31.6%; Pred. No. 1,2e-12;
Matches 85; Conservative 44; Mismatches 120; Indels 20; Gaps 11;

QY 369 QCLISDSGVLLBSNLIKVLPTWSTPYEPKSCDKHTHTCPCPAP--ELLGG--PSVFLP 423
DB 20 QCLGQKSAACHVEYNSVI--NESLPVPFPDCCPANSCTCPSSSRRLISGQPSLSLQR 77
QY 424 PKPDTLMISRTPEVTCVVDVSHDEDEYKFMNVYDGEVHNNAKTKPREQYNSTYRVVS 483
DB 78 PDLGD-LLLGRDASLTCTLSGLKNPDAV-FTW--EFTNGNEPVQQAQRDLSCGYSVSS 133
QY 484 VLTVHMDWLNKREYKCVSNKALPA-PIEKTISSAKGQPREPOVYTLPPSRDELTKN-Q 541
DB 134 VLPSAETWAKARIEFTTVTHPEIDSGSLTATISRGVTF--FQVHLPPREELALNEQ 191
QY 542 VSLTCLVKGFPSPDIAYEWESNQ--PENNY---KTPPVLDSDGSFFLYSKLTVDKSRW 596
DB 192 VTLTCLVGRFSPKQVLSVMHQEQEVEDSFLVWKSMPSSQDKATYATISLRLVPAEDW 251
QY 597 QQGVNFSQSVMEALHNHYTKSLSPG 625
DB 252 NQGDVYSQWVHGELAEHFTQKTTIDRLAG 280

RESULT 47
MUC_CHICK STANDARD; PRT; 446 AA.
ID 21-JUL-1986 (Rel. 01, Created)
AC P01875;
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DB Ig mu chain C region.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
RN SEQUENCE OF 80-446 FROM N.A.
RP MEDLINE=83299221; Pubmed=6310496;

```



RA Dahan A., Reynaud C.-A., Weill J.-C.;  
 RT "Nucleotide sequence of the constant region of a chicken mu heavy  
 chain immunoglobulin mRNA."  
 RL Nucleic Acids Res. 11:5381-5389(1983).  
 RN [2]  
 RP SEQUENCE OF 1-79 FROM N.A.  
 RA Weill J.-C.;  
 RL Submitted (MAR-1986) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X01613; CAA25762.1; -  
 DR PIR; A02170; MHCH.  
 DR HSP; P01857; 1FC1.  
 DR InterPro; IPR007110; Ig\_1like.  
 DR InterPro; IPR003597; Ig\_c1.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00047; Ig\_3.  
 DR SMART; SM00407; IgC1; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 DR Immunoglobulin domain; Immunoglobulin C region.  
 KM NON\_TER 1 1  
 FT DOMAIN 1 105 CH1.  
 FT DOMAIN 106 209 CH2.  
 FT DOMAIN 210 316 CH3.  
 FT DOMAIN 317 427 CH4.  
 FT DOMAIN 428 446 C-TERMINAL REGION.  
 FT DISULFID 16 16 INTERCHAIN (WITH A LIGHT CHAIN)  
 FT DISULFID 27 85 (BY SIMILARITY).  
 FT DISULFID 135 190 BY SIMILARITY.  
 FT DISULFID 237 266 BY SIMILARITY.  
 FT DISULFID 284 284 INTERCHAIN (WITH A HEAVY CHAIN)  
 FT DISULFID 344 406 (BY SIMILARITY).  
 FT DISULFID 445 445 INTERCHAIN (WITH A HEAVY CHAIN)  
 FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 446 AA; 48173 MW; 3CB0CE108949BD17 CRC64;  
 Query Match 8.7%; Score 298; DB 1; Length 446;  
 Best Local Similarity 27.4%; Pred. No. 5.8e-12;  
 Matches 115; Conservative 62; Mismatches 175; Indels 66; Gaps 21;  
 QY 247 SSKSWTTPFLKNEVSVKAVTODPKLQMGKPLHLTLPOALPOYAGSNLTALAEAKTG 306  
 DB 35 SSIAPFDSNNSSVSGMDVI--PKVISG-----PPYAVSR--IQMNSSEG 77  
 QY 307 KLHGEVNLVVRATOLQKLTCEVMP--TSRK--LMLSLKLENK----- 348  
 DB 78 KEKOPR---CRAPRGRNVEVSVNPGPIPTNGPLPLVTMPPSRKEDEGPFNNASTL 134  
 QY 349 AKVSKREK---VWVLN--PEAGMOCLLSDSGVLLSENIKVLPT--NSTPV----- 394  
 DB 135 CQTRGRARRPTEVWYNGSPVAATAATATTVGEVVAESRIISTSEEMTGTATFSCVNG 194  
 QY 395 EPKSCDKTTC--PPCPABELLGGSEVFLPPPKKDTLMISTREPTCVVVDVSHEDP-E 451  
 DB 195 EKRNTSKRMKCEGLPEVVODI-----AIRVITSEFVD-IFISKATLTCVSNMNVNADGLE 249

QY 452 VKFNMYVDGVEVYNAKTRPREQVNSTYRVVSVLTFLHQDMLNGKEYCKVSNKALPAPI 511  
 DB 250 VSM-WKEGKGKLETLGK-RVLQSGNGLTYVSVAVLVCASFNDDGGGVCYKNHPLLEPFM 307  
 QY 512 EKTISKAK-GQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVWESNQG--PE 567  
 DB 308 EKKMRKTSANRPPSVYVFPPTPEQLNGNQLSTVTCMAQGNPPLLFRRMRNGEPLPQ 367  
 QY 568 NNYKTPPLDLS--DGSFVLVSKLTVDSRWQGVFSCSVWHEALHNHYTQKSLSLSPG 625  
 DB 368 SOSVTSAPVAENPENESVYVAVSLVGABEWAGNVYTCLVGHEALPLQLAQKSDRASG 427  
 RESULT 48  
 HVC2\_HETFR  
 ID HVC2\_HETFR STANDARD; PRT; 438 AA.  
 AC P23085;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig heavy chain C region (Clone 12022) (Fragment).  
 OS Heterodontus francisci (Horn Shark).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;  
 OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;  
 OC Heterodontidae; Heterodontus.  
 OX NCBI\_TaxID=7792;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=88328985; PubMed=3138109;  
 RX Kokubu F., Hinds K., Liltman R., Shambloet M.J., Liltman G.W.;  
 RT "Complete structure and organization of immunoglobulin heavy chain  
 constant region genes in a phylogenetically primitive vertebrate."  
 RL EMBO J. 7:1979-1988(1988).  
 CC -1- SIMILARITY: Contains 4 immunoglobulin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X07784; CAA30617.1; -  
 DR PIR; S00980; HYRKC2.  
 DR HSP; P01842; 7FAB.  
 DR InterPro; IPR007110; Ig\_1like.  
 DR InterPro; IPR003597; Ig\_c1.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00047; Ig\_4.  
 DR SMART; SM00407; IgC1; 2.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.  
 KM NON\_TER 1 1  
 FT DOMAIN 1 101 IG-LIKE 1.  
 FT DOMAIN 108 202 IG-LIKE 2.  
 FT DOMAIN 213 305 IG-LIKE 3.  
 FT DOMAIN 315 415 IG-LIKE 4.  
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 438 AA; 47904 MW; 6E67A9C5BBF7AA7B CRC64;  
 Query Match 8.7%; Score 297; DB 1; Length 438;  
 Best Local Similarity 27.8%; Pred. No. 6.5e-12;  
 Matches 105; Conservative 59; Mismatches 148; Indels 66; Gaps 18;



[illegible]

RL EMBO J. 7:1979-1988(1988).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

.....

DR EMBL: X07781: CAA30614.1: ALT SEQ.

PTB: S01854: HVBKCO

DA 12K, 501004, 117KACU;  
HSCD: 001043- 75AB  
DB

DK HSBF; EUL842; /FAB.

DK InterPro; IPR007110; Ig-like.

DR InterPro; IPR003597; Ig\_c1.

DR InterPro; IPR003006; Ig\_MHC.

DR Pfam: PF00047: iq: 4.

DR SMART: SM00407: IGcl: 2.

PROCTER, BSE0825. IQ TYPE, A

DA FNOV01E, F030053, 10\_111E, 4.  
DE PROJECT, P000000 10\_111E 3

DK PROSIE; PS00290; LG\_MHC; 3.

KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;

KW Alternative splicing; Transmembrane.

FT	NON	TER	1	1
----	-----	-----	---	---

FT	DOMAİN	1	99	CH1.
----	--------	---	----	------

FT	DOMAIN	100	205	CH2.
FT	DOMAIN	206	308	CH3.
FT	DOMAIN	309	418	CH4.
FT	TRANSMEM	438	458	POTENTIAL.
FT	CARBOHYD	164	164	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	200	200	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	245	245	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	275	275	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	374	374	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	411	411	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	415	415	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	461 AA;	50762 MM;	029EE91C8805DF911 CRC64;

## RESULT 50

## HVCS HETER

IN HVCS HETEP STANDARD: DPT: 438 AA

AD	010700Z	FM
NO	03067.	

AC F23087; 2000 (5-1-2000)

01-NOV-1991 (REL. 20, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Id heavy chain C region, secreted form (C1one 3050).

*Heterodontus francisci* (Horn shark).

Phylum: Chordata: Vertebrata: Chondrichthyes: Eulacynota: Metazoa: Cnidaria: Vertebrata: Chondrichthyes:

[illegible]

Elasmobranchii; Galeomorphii; Heterodontiformes; Heterodontiformes;

OC Heterodontidae; Heterodontus.

0X NCBI\_TaxID=7792;

$$\mathbf{R}^N \quad [1]$$

RP SEQUENCE FROM N.A.

```

TTSITE=$n]een:

```

PY MEDLINE+0803089E; PubMed-3130106;

```
KA MEDLINE=80040903; PubMed=3130107;  
rcubd n m d g b h l t w r f a c w
```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X07761; CAA30613.1; ALT\_INIT.

DR PIR; S01853; HYRKS.

DR HSRP; P01842; 7FAB.

DR InterPro; IPR007110; Ig-1like.

DR InterPro; IPR003597; Ig\_c1.

DR InterPro; IPR003006; Ig\_MHC.

DR Pfam; PF00047; Ig; 4.

DR SMART; SM00407; IGc1; 2.

DR PROSITE; PS50835; IG\_LIKE; 4.

DR PROSITE; PS00290; IG\_MHC; 3.

KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;

KW Alternative splicing.

FT NON\_TER 1 1

FT DOMAIN 1 99

FT DOMAIN 100 205

FT DOMAIN 206 308

FT DOMAIN 309 418

FT CARBOHYD 164 164

FT CARBOHYD 200 200

FT CARBOHYD 245 245

FT CARBOHYD 275 275

FT CARBOHYD 374 374

FT CARBOHYD 411 411

FT CARBOHYD 415 415

FT CARBOHYD 425 425

SEQUENCE 438 AA; 47920 MW; 9BDF37C3BD7D15B CRC64;

Query Match

Best Local Similarity 25.6%; Score 293.5; DB 1; Length 438;

Matches 117; Conservative 66; Mismatches 163; Indels 105; Gaps 22;

208 SSVIVKKEGQVEF---SFLPFT-----VEKLTGSGELMQAERASSKSW 251

37 TSVSWKKDNEPIITGLKTPSVLNKKGTYSQSLTITSEVSSSKICEVRGES--VW 94

252 IFFDLKNEKESVVRVTDPKLQMGKRLPLHLTLPOAYAGSGNTLALAE----- 303

95 I-----KEI-----PDCK---GDKVHPVILITSSSEITSRPFTVLCSTIDFHPES 139

304 -----KTGKLHQEVALVVMRAITQLOKNLTCEVWGPTSPKMLSLKLENKAIVSKREK 357

140 ITVSWMKDQO-HHESGFVT-----SPTCGVNGTFSATSLTL-----VPAPE-- 179

358 VVVLNDEAGMMQCLSDSGQVLIESNIKVLPTWSTVEPKSCDHTTCCPPAPELLGSP 417

180 -HFTNK---VYTCQVSHOG-VTQSRNI-----TSQVPCSCND----P 213

418 SVFLPRPKKDKDMLMISRTPEVTCVVVDVSHEDPEVKFNNYVVDGVENHAKTKRREKQVNS 477

214 VIKLPPSTIEQVLT-LEATVTLTCTV---VSNAPRGVAVSW---TQEQSKLSKSEIAYVQGED 266

478 TVRVVSVLTVLHQDMLNGKEKVKVSKVKNALPARIETKISKAGQP--REPOV--YTLPPSPD 535

267 ADSVISTVNIISQAWLSGAEFYCVVNHODLPPLRLASIKHEKVDLRSEVSILSLSPAD 326

536 ELTKQVSLTCLVKGFFPSDIAVENSQOPEN--NYKTTPLVDS--GSFPLYSKLTVD 592

327 VSAQRFLSLTCLVRGFFPREIFVKMTVNDKSVNPGNYKNTVEVAENDNSYFIYSLSLIA 386

593 KSRMOGNNFSCSVMEALHNHTQKSLSPG 625

387 ABEWASGASYSVGVGHEALPLKINTVKKSSG 419

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

RESULT 51

ALC2 HUMAN  
ID ALC2 HUMAN STANDARD; PRT; 340 AA.  
AC P01877;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE IG alpha-2 chain C region.  
GN IGH2.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=84130179; PubMed=6421489;

RA Flanagan J.G., Lefranc M.-P., Rabbitts T.H.;

RT "Mechanisms of divergence and convergence of the human immunoglobulin

alpha 1 and alpha 2 constant region gene sequences.";

RL Cell 36:681-688(1984).

RN (2)

RP SEQUENCE (BUT).

RX MEDLINE=78137069; PubMed=416441;

RA Torano A., Putnam F.W.;

RT "Complete amino acid sequence of the alpha 2 heavy chain of a human

IgA2 immunoglobulin of the A2m (2) allotype.";

RL Proc. Natl. Acad. Sci. U.S.A. 75:966-969(1978).

RN (3)

RP SEQUENCE (MYELOMA PROTEIN LAN).

RX MEDLINE=79180140; PubMed=286295;

RA Tsuzukida Y., Wang C.-C., Putnam F.W.;

RT "Structure of the A2m (1) allotype of human IgA -- a recombinant

molecule.";

RL Proc. Natl. Acad. Sci. U.S.A. 76:1104-1108(1979).

RN (4)

RP REVIEW.

RX MEDLINE=91054387; PubMed=2241915;

RA Kerr M.A.;

RT "The structure and function of human IgA.";

RL Biochem. J. 271:285-296(1990).

CC -I- FUNCTION: Ig alpha is the major immunoglobulin class in body

secretions. It may serve both to defend against local infection

and to prevent access of foreign antigens to the general

immunologic system.

CC -I- SUBUNIT: Monomeric or polymeric.

CC -I- MISCELLANEOUS: The sequence of the A2m(1) allotype is shown.

CC -I- SIMILARITY: Contains 3 immunoglobulin-like domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; J00221; AAB59396.1; ALT\_INIT.

DR PIR; B22360; B22360.

DR HSRP; P01810; 2FBU.

DR Genew; HGNC:5479; IGH2.

DR MIM; 147000; --.

DR GO; GO:0005624; C:membrane fraction; NAS.

DR GO; GO:0001823; F:antigen binding; TAS.

DR GO; GO:0006955; F:immune response; NAS.

DR InterPro; IPR007110; Ig-1like.

DR InterPro; IPR003597; Ig\_c1.

DR InterPro; IPR003006; Ig\_MHC.

DR Pfam; PF00047; Ig; 3.

DR SMART; SM00407; IGc1; 2.

DR PROSITE; PS50835; IG\_LIKE; 3.

DR PROSITE; PS00290; IG\_MHC; 2.

KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.

FT NON\_TER 1 1

FT DOMAIN 6 98

IG-LIKE 1.



```

FT DOMAIN 112 207 IG-LIKE 2.
FT DOMAIN 215 317 IG-LIKE 3.
FT DISULFID 26 85 PROBABLE.
FT DISULFID 101 101 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 109 109 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 110 107 PROBABLE.
FT DISULFID 134 191 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 169 169 INTERCHAIN (WITH HEAVY CHAIN IN ANOTHER
FT DISULFID 179 179 SUBUNIT) (POTENTIAL).
FT DISULFID 237 300 PROBABLE.
FT DISULFID 339 339 INTERCHAIN (WITH J CHAIN) (PROBABLE).
FT CARBOHYD 47 437 N-LINKED (GLCNAC. . .).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .).
FT CARBOHYD 131 131 N-LINKED (GLCNAC. . .).
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .).
FT CARBOHYD 327 327 P -> S (IN A2M(2) ALLOTYPE).
FT VARIANT 93 93 P -> S (IN A2M(2) ALLOTYPE).
FT VARIANT 102 102 P -> R (IN A2M(2) ALLOTYPE).
FT VARIANT 279 279 F -> Y (IN A2M(2) ALLOTYPE).
FT VARIANT 296 296 D -> E (IN A2M(2) ALLOTYPE).
FT VARIANT 326 326 V -> I (IN A2M(2) ALLOTYPE).
FT VARIANT 335 335 V -> A (IN A2M(2) ALLOTYPE).
SQ SEQUENCE 340 AA; 36508 MW; 9892270756F3276 CRC64;

```

```

Query Match 8.3%; Score 282.5; DB 1; Length 340;
Best Local Similarity 28.3%; Pred. No. 4e-11;
Matches 100; Conservative 47; Mismatches 137; Indels 69; Gaps 17;

Qy 333 PTPSPKLM-LSLKLENKAKVSKR-----EKPVNLNPEAGM-----MOCILSDSG 377
Db 3 PTPSPKLPFLSDSTPQDNVAVACLVQGFPOEPLSVTWSGQVNTARNPPPSODASGD 62
Qy 378 VLLESNIKVLPTWSTPVEPKS--CD-KHTHCP-----PCPAPELLGSPVFLPPPK-- 426
Db 63 LYTTSSQLTPATPOCP-DGKSVTCHVKATYNSQDVTYPCFP-----PPPPCC 110
Qy 427 -----KDTLMISRTPEVTCVVDVSHEDPEVKENYVDGVEVNAKTKPREQYN 476
Db 111 HPRLSIHRPALEDLLIGSEANLTCTLGL-RDASGATITWPPSSCK--SAVGPPERDLC 167
Qy 477 STYRVSVLTVLHODWLNKGEYKCKVSKALPAPIEKTISAKGQPREPOVYTLPPSKDE 536
Db 168 GCYSVSVTLPCGCAOPWNGHGETFTCTAHPBELKTPLTANITKS-GNTPRPVHLLPPPSEE 226
Qy 537 LTKNO-VSLTCLVKGFPYSDIAVENESNGQ--PENNYTTPPYLD-SDG--SFLYSKLT 590
Db 227 LALNELVLTCLARGFSPKDVLRWLGQSQELPREKVTLMASROBPSGTTTFATSTLR 286
Qy 591 VDKSRMOQGNVFSQVMEHALHNHYTOKSLSPS-----LQIDETC 632
Db 287 VAAEDWKKGGTFSSCMVGHGHALPLAFTQKTIRLAKGPRHVVVSVMAEVDTC 339

```

```

RESULT 52
HVC3_HETFR STANDARD; PRT; 393 AA.
AC P23086;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DB Ig heavy chain C region (Clone 6121) (Fragment).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
NCBI_TaxID=7792;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=6832985; PubMed=1138109;
RA Kokubo F., Hinde K., Litman R., Shambloet M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
RL EMOB J. 7:1979-1988(1988).
CC -I- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
CC EMBL; X07782; CAA30615.1; -.
CC PIR; S01852; HVRKCL.
CC HSSP; P01842; 7PAB.
CC InterPro; IPR007110; Ig_Like.
CC InterPro; IPR003597; Ig_C1.
CC InterPro; IPR03006; Ig_MHC.
CC Pfam; PF00047; Ig_3.
CC SMART; SM00407; IgC1_2.
CC PROSITE; PS00835; IG_LIKE_3.
CC PROSITE; PS00290; IG_MHC_3.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON_TER 1
FT DOMAIN 1 157 IG-LIKE 1.
FT DOMAIN 168 260 IG-LIKE 2.
FT DOMAIN 270 370 IG-LIKE 3.
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 393 AA; 43081 MW; 4E4AD076972F18B5 CRC64;

```

```

Query Match 8.2%; Score 281.5; DB 1; Length 393;
Best Local Similarity 26.7%; Pred. No. 5.6e-11;
Matches 116; Conservative 63; Mismatches 164; Indels 91; Gaps 22;

```

```

Qy 211 VYKKEGQVSPPLAFTVEKLTGSGELMWQAEBSKSWITFDLKNKEYSVKRTQDP 270
Db 13 VLNKKGYTYTQ-SSQLTIT-ESEVSSKIYCEVARGES--VMI-----KEI-----PDC 56
Qy 271 KLGQKKLPLHLTLPLQALPGYAGSNLTALAE-----KTGLHGVNLV 316
Db 57 K--GDVNHPTVLTITQSSSEITSRPRTVLCSTIIDHPISITVWLDQO-HHESGFT 112
Qy 317 MRATOLKNLTCEWGFSPSKMLSLKLENKAKVSKREKRPVWVNLNPEAGMOCILSDSG 376
Db 113 -----SPTCGVNGFSATSNLT-----VPARE--WPTNK---YITGVSHQG 149
Qy 377 QVLESNIKVLPTWSTPVEPKSCDKHTTCCPAPPELLGSPVFLPPPKKDTLMISRT 436
Db 150 -VTQSRNI-----TSGVQPCSCND---PVKILPPSIEQVL-LEATV 186
Qy 437 EVTCVVDVSHEDPEVKENYVDGVEVNAKTKPREQYNSTYRVSVLTVLHODWLNK 496
Db 187 TLNLT--VSNAIPGVNSW--TQEQSKLSKSEIAVPGEDADSVISTVNSTQAMLSGA 240
Qy 497 EYCKVSNKALPAPIEKTISKAGQV-REPQV-YTLPPSRDELTKNOVSLTCLVKGFP 554
Db 241 EFTCVVNHQDLPPTPLRSIHKEVYKDLREBSVSLTSPADVSNQRSLTCLVKGFSR 300
Qy 555 DIAVESNGQPEN--NYKTTTPVLDSD-GSFLYSKLTVDKSRMOQGNVFSQVMEHAL 611

```



Db 301 ELPVKTINDKSVNPGNYKNTVEAMENDNSYFYISLLSLIAAEWASGASVCVGHENI 360  
OY 612 HHNYTOKSLSPG 625  
Db 361 PLKINRTVYKSSG 374

RESULT 53  
ALCI HUMAN  
ID ALCI HUMAN STANDARD; PRT; 353 AA.  
AC P01876;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Ig alpha-1 chain C region.  
GN IGHAI.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84130179; PubMed=6421489;  
RA Flanagan J.G., Lettranc M.-P., Rabbitts T.H.;  
RT "Mechanisms of divergence and convergence of the human immunoglobulin  
RT alpha 1 and alpha 2 constant region gene sequences.";  
RL Cell 36:681-688(1984).  
RN [2]  
RP SEQUENCE (MYELOMA PROTEIN BUR), AND DISULFIDE BONDS.  
RX MEDLINE=79151016; PubMed=107164;  
RA Putnam F.W., Liu Y.-S.V., Low T.L.K.;  
RT "Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal  
RT IgA1 protease, digestion, Fab and Fc fragments, and the complete  
RT amino acid sequence of the alpha 1 heavy chain.";  
RL J. Biol. Chem. 254:2865-2874(1979).  
RN [3]  
RP SEQUENCE (MYELOMA PROTEIN TRO).  
RX MEDLINE=76023781; PubMed=809331;  
RA Kratzin H., Altevogt P., Ruban E., Kortt A., Staroscek K.,  
RA Hilschmann N.;  
RT "The primary structure of a monoclonal IgA-immunoglobulin (IgA Tro.),  
RT II. The amino acid sequence of the H-chain, alpha-type, subgroup III;  
RT structure of the complete IgA-molecule.";  
RL Hoppe-Seyler's Z. Physiol. Chem. 356:1337-1342(1975).  
RN [4]  
RP DISULFIDE BONDS.  
RX MEDLINE=80114124; PubMed=393607;  
RA Yang C.-Y., Kratzin H., Gotz H., Hilschmann N.;  
RT "Rule of antibody structure. Primary structure of a human monoclonal  
RT IgA-immunoglobulin (myeloma protein Tro). VII. Purification and  
RT characterization of the disulfide bridges.";  
RL Hoppe-Seyler's Z. Physiol. Chem. 360:1919-1940(1979).  
RN [5]  
RP REVIEW.  
RX MEDLINE=91054387; PubMed=2241915;  
RA Kerr M.A.;  
RT "The structure and function of human IgA.";  
RL Biochem. J. 271:285-296(1990).  
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body  
CC secretions. It may serve both to defend against local infection  
CC and to prevent access of foreign antigens to the general  
CC immunologic system.  
CC -1- SUBUNIT: Monomeric or polymeric.  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: J00220; AAC82528.1; ALT\_INIT.  
DR PIR: A22360; AIHU.  
DR PDB: 1IGA; 15-JUN-99.  
DR Genew; HGNC:5478; IGHAI.  
DR MIM; 146900; -.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006955; P:immune response; NAS.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003597; IG cl.  
DR InterPro; IPR003006; IG\_MHC.  
DR Pfam; PF00047; Ig; 3.  
DR SMART; SM00407; IG1; 2.  
DR PROSITE; PS00835; IG LIKE; 3.  
DR PROSITE; PS00290; IG\_MHC; 1.  
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;  
KW Polymorphism; 3D-structure; Repeat.  
FT NON\_TER 1 1  
FT DOMAIN 6 98  
FT FT 125 220  
FT DOMAIN 228 330  
FT FT 14 14  
FT DISULFID 26 85  
FT FT 77 101  
FT DISULFID 122 122  
FT DISULFID 123 180  
FT DISULFID 147 204  
FT FT 182 182  
FT DISULFID 192 192  
FT FT 250 313  
FT DISULFID 352 352  
FT CARBOHYD 105 105  
FT CARBOHYD 111 111  
FT CARBOHYD 113 113  
FT CARBOHYD 119 119  
FT CARBOHYD 121 121  
FT CARBOHYD 144 144  
FT CARBOHYD 340 340  
FT VARIANT 176 176  
FT FT 163 165  
FT CONFLICT 176 176  
FT CONFLICT 190 190  
FT CONFLICT 227 227  
FT CONFLICT 231 231  
FT CONFLICT 290 290  
SQ SEQUENCE 353 AA; 37654 MW; EBA11ECB7E85DB21 CRC64;

Query Match 8.2%; Score 280.5; DB 1; Length 353;  
Best Local Similarity 26.8%; Pred. No. 5.6e-11;  
Matches 93; Conservative 49; Mismatches 132; Indels 73; Gaps 13;

OY 305 TGKLGQEVNLVVMARAT--LQKNLTCEVNGVPTSPKLMSTLKENKAQVSRKRPVWVLN 362  
Db 60 SSDLTYTSSQULTPATOCLAGKSVTCHVXHTNPSQDVTVP----- 100  
OY 363 PEAGMGOCLDSQGVLESNIKVLPTWSTPEVPSKCDKHTCPCPAPBELLGGPSVFL 422  
Db 101 -----CPVSTPTPT-----SSTPTPTPSGCC-HPRLSLH 130  
OY 423 PPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNNYVDSGEVFNAAKTKREQDYNSTRYV 482  
Db 131 RPALEED-LLTSEANLTCTLTGL-RDASGVFTFTWPSGK--SAVQGPERRDLGGCVSVS 186  
OY 483 SVLTATVHODMNGKGEYKCKVSNKALPAPIETGISAKAQPREPOVYTLPPSRDELTKNQ- 541  
Db 187 SVLPCCAPPMWNGKFTTAAIYPSKTPILTYSKS-GNTRPPEVHLPPSEELALNEL 245  
OY 542 VSLTGLVGFYPSDIAVWESNGQ--PENNYKTPPVLD-SDG--SFFLYSKLTIVDKSRW 596  
Db 246 VTLTGLANGFSPKQVLAIVWLGSGQDLPREKVLTAASRQEPGQTTTFAVTSILVAEDW 305



```

QY 597 OOGNFSCSVHHEALHNHYTOKSLSLSPG-----LQDSTC 632
DB 306 KKGDTFSCVGHLEALPLAFTOKTIDRLAGKPTHVNVSVMAEVDGTC 352

RESULT 54
ALC1_GORGO STANDARD; PRT; 353 AA.
ID ALC1_GORGO
AC P20758;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig alpha-1 chain C region.
GN IGHAI.
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Hominoidea; Gorilla.
OX NCBI_TaxId=9595;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph node; PubMed=2506527;
RX MEDLINE=89386006;
RA Kawamura S., Omoto K., Ueda S.;
RT "Nucleotide sequence of the gorilla immunoglobulin alpha 1 gene.";
RL Nucleic Acids Res. 17:6732-6732(1989).
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body
CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15045; CAA33147.1; ALT_INIT.
DR HSSP; P01810; 2PB1.
DR InterPro; IPR007110; Ig_1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON_TER 1
FT DOMAIN 6 98
FT DOMAIN 125 220
FT DOMAIN 228 330
FT DISULFID 14 14
FT DISULFID 26 85
FT DISULFID 77 101
FT DISULFID 122 122
FT DISULFID 123 180
FT DISULFID 147 204
FT DISULFID 182 182
FT DISULFID 192 192
FT DISULFID 250 313
FT DISULFID 352 352
FT CARBOHYD 144 144
FT CARBOHYD 340 340
FT SEQUENCE 353 AA; 37755 MW; 482088DB02AC7514 CMC64;

Query Match 8.1%; Score 278; DB 1; Length 353;
Best Local Similarity 28.1%; Pred. No. 8.1e-11;

```

```

Matches 101; Conservative 47; Mismatches 143; Indels 68; Gaps 17;
QY 333 PTSEKLM-LSLKLNKAKVSKR-----EKPVWLNDEAGM-----WQCLLSDSQ 377
DB 3 PTSPKVFPLSLCSTQPDGDVVAACLVGFPQDEPLSLTWSESGGVTAARPPESQDASGD 62
QY 378 VLBSNLIKVLPTWSTPYEPKS--CDKTH-----TCPPCPAPELIGGSVF 420
DB 63 LYTTSSQTLPLATQCP-DGKSVTCHVNHYNPSSQDVTVPGRVSTPPTPSPSTPPTS-- 119
QY 421 LFPF-----KPK-DTLMISRTPEVTCVVVDVSHEDDEVKRWVVDGVENNAATKP 470
DB 120 --PPCCHPLSLHRLPALEDLILGSEANLTCTLTGL-RDASGVFTWTPSSGK--SAVECP 174
QY 471 REEQYNSTYRVVSVLTVLHODWLNGEKVKVSKALPAPIEKTISKAKQPREQVYTL 530
DB 175 PERLCCGCVSVSLTPCCAPPMHNGKFTCTAAVPEKTLTLTLSS--GNMPPPEVHL 233
QY 531 PPSRDELTKQV-VSLTGVKGFYPSDIAVEMESNQ--PENNYKTPPVLD-SDG--SFF 584
DB 234 PPSSEELALNELVTLTCLARFSPKDVLYRWLGSSQELPREKYLTMASRQEPSSQTTTFA 293
QY 585 LYSKLTVDKSRMOOGNFSCSVHHEALHNHYTOKSLSLSPG-----LQDSTC 632
DB 294 VTSILRVAAEDWKXKGDTFSCVGHLEALPLAFTOKTIDRLAGKPTHVNVSVMAEVDGTC 352

RESULT 55
HVC1_HETFR
ID HVC1_HETFR STANDARD; PRT; 370 AA.
AC P23084;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain C region (Clone 6125) (Fragment).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxId=7792;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=88328985; PubMed=3138109;
RX Kokubu F., Hinds K., Litman R., Shambloot M.J., Litman G.W.;
RA "Complete structure and organization of immunoglobulin heavy chain
RT constant region genes in a phylogenetically primitive vertebrate.";
RL EMBO J. 7:1979-1988(1988).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X07783; CAA30616.1; -.
DR PIR; S01851; HVKRC5.
DR HSSP; P01857; 1FC1.
DR InterPro; IPR007110; Ig_1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 3.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON_TER 1
FT DOMAIN 40 134
FT DOMAIN 145 237
FT DOMAIN 247 347
FT CARBOHYD 98 98
FT SEQUENCE 370 AA; 37755 MW; 482088DB02AC7514 CMC64;

Query Match 8.1%; Score 278; DB 1; Length 353;
Best Local Similarity 28.1%; Pred. No. 8.1e-11;

```







```

RL Gene 13:365-374(1981).
RN [2]
RP SEQUENCE OF 1-213 (MOPC 47A).
RX MEDLINE=80049769; PubMed=115869;
RA Robinson E.A., Appella E.;
RT "Amino acid sequence of a mouse myeloma immunoglobulin heavy chain
  (MOPC 47 A) with a 100-residue deletion.";
RN J. Biol. Chem. 254:11418-11430(1979).
RP [3]
RX SEQUENCE OF 1-254 AND 291-344 (M511).
RA MEDLINE=81054880; PubMed=6776528;
RA Robinson E.A., Appella E.;
RT "Complete amino acid sequence of a mouse immunoglobulin alpha chain
  (MOPC 511).";
RN Proc. Natl. Acad. Sci. U.S.A. 77:4909-4913(1980).
RL - FUNCTION: Ig alpha is the major immunoglobulin class in body
  secretions. It may serve both to defend against local infection
  and to prevent access of foreign antigens to the general
  immunologic system.
CC - MISCELLANEOUS: The final C-region domain is deleted from Ref.2
  chain. It was isolated from a myeloma protein that contains 1
  light and 1 heavy chain per molecule, linked by a disulfide bond.
  In contrast, normal mouse IgA molecules contain 2 light and 2
  heavy chains and lack a light-heavy chain disulfide bond.
CC - MISCELLANEOUS: M511 chain was isolated from a myeloma protein that
  binds phosphorylcholine.
CC - MISCELLANEOUS: M511 sequence was compared with that of mouse MOPC
  47A, and a genetic mechanism for the deletion of the CH3 domain of
  the mutant chain is proposed.
CC - SIMILARITY: Contains 3 immunoglobulin-like domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
DR EMBL, D11468; BA002026.1; ALT_INT.
DR PIR, A91479; AHMS.
DR HSSP, P01810; 2PBJ.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON TER 1 1
FT DOMAIN 6 99
FT 116 206 IG-LIKE 1.
FT DOMAIN 219 321 IG-LIKE 3.
FT DISULFID 26 84 BY SIMILARITY.
FT DISULFID 76 100 BY SIMILARITY.
FT DISULFID 114 171 BY SIMILARITY.
FT DISULFID 138 195 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .).
FT CARBOHYD 101 101 O-LINKED.
FT CARBOHYD 331 331 O-LINKED.
FT VARIANT 2 2 S -> A (IN MOPC 47A AND M511).
FT VARIANT 18 18 S -> C (IN MOPC 47A AND M511).
FT VARIANT 67 67 N -> S (IN MOPC 47A AND M511).
FT VARIANT 73 73 A -> T (IN MOPC 47A).
FT VARIANT 112 112 P -> G (IN M511).
FT VARIANT 135 135 S -> Q (IN MOPC 47A AND M511).
FT VARIANT 141 141 N -> D (IN MOPC 47A AND M511).
FT VARIANT 168 168 Q -> E (IN MOPC 47A).
FT VARIANT 212 213 VT -> SQ (IN MOPC 47A).
FT VARIANT 235 235 E -> G (IN M511).
FT VARIANT 255 290 MISSING (IN M511).
FT VARIANT 295 295 T -> D (IN M511).

```

```

FT VARIANT 301 301 Q -> G (IN M511).
FT VARIANT 329 329 N -> Q (IN M511).
FT VARIANT 331 331 S -> N (IN M511).
SQ SEQUENCE 344 AA; 36876 MW; 3694CF99B19A9F8 CRC64;
Query Match 7.3%; Score 248.5; DB 1; Length 344;
Best Local Similarity 25.5%; Pred. No. 6.1e-09;
Matches 93; Conservative 56; Mismatches 145; Indels 71; Gaps 14;
QY 282 LTLPQALP-----QYASGNLTLLAEKGTGLHDEVNVLVYMRALQOLNSLTCEV 330
  |||||
  11 LTPPALSSDPVIGCLINDYFPSPGTNNVTW-GKSGDITVTVPFPPALASGGRYTMNQL 69
QY 331 WGPSPFLMTSLKLENKAIVSKREKVPWLNPEAGMOCILSDSGVLLSNIKVL-PT 389
  |||||
  70 TLPA-----VECBGEVSKSVQVHDSNPVELDYNCGPT 104
QY 390 WSPVPEKSCDKHTCPCPAPELLGSPVFLPPPKKDTLMSRPEVTGVVDVSHED 449
  |||||
  105 PPPPI-----TIPSCQ-----PSLSQRPALPD-LILGSDASITCTLNGLRNPE 147
QY 450 PEYKFNMYVDGVYHNAKTPREEQNST---YRVSVLTVLHODMLNGEYKCKNSKA 506
  |||||
  148 GAV-FTW-----EPTGCKDAVQKKAQVNSGCVSVSLGCAERMNSGASFCKTVTHPE 201
QY 507 LPAPIEKTIKAKQGPPEPVVTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVEMESNGQ 565
  |||||
  202 -SGLTGTITAKVYVNTTPPVVHLPPSEELANELLSTCLVRAFPKVKVLRMLHGNE 260
QY 566 ---PENNYKTPPYLSDG--SFFLYSKLTVDKSRMOQGNVFCSGVMEALHNHYTKSL 620
  |||||
  261 ELSPESYLVEPELPKEPEGATVTLVTSVLSAETWKGQDQYSCMVGHSLPMNFTQXTI 320
QY 621 SLSPG 625
DB 321 DRUSG 325

```

## RESULT 58

```

SHS1_MOUSE STANDARD; PRT; 513 AA.
ID SHS1_MOUSE
AC P97797; O08907; O35924; O88555; O88556; P97796; Q8R559; Q9QX57;
AC Q9WTN4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein-tyrosine phosphatase non-receptor type substructure 1 precursor
DE (SHP substrate-1) (SHPs-1) (inhibitory receptor SHPS-1) (Signal-
DE regulatory protein alpha-1) (SIRP-alpha-1) (msIRP-alpha1) (MylD-1
DE antigen) (Brain Ig-1-like molecule with tyrosine-based activation
DE motifs) (Blt) (p84).
GN PPNS1 OR SHPS1 OR SIRP OR MYD1 OR BIT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain;
RX MEDLINE=97223399; PubMed=9070220;
RA Yamao T., Matozaki T., Amano K., Matsuda Y., Takahashi N., Ochi F.,
RA Fujioke Y., Kasuga M.;
RT "Mouse and human SHPS-1: molecular cloning of cDNAs and chromosomal
RT localization of genes.";
RL Biochem. Biophys. Res. Commun. 231:61-67(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS ALA-29; ARG-67; ARG-91;
RP THR-96; SER-128; PRO-194 AND ASN-224.
RC STRAIN=BA1B/C; TISSUE=Brain;
RC MEDLINE=97230468; PubMed=9073522;
RA Ohnishi H., Kubota M., Sano S.-I.;
RT "Blt (Blt) maps to mouse chromosome 2.";
RL Genomics 40:504-506(1997).

```



```

RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), SEQUENCE OF 32-53 AND 422-433
RP (ISOFORM 2), VARIANTS ALA-29; ARG-91; THR-96 AND SER-128,
RP N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
RX STRAIN=BALB/c; TISSUE=Brain, and Cerebellum;
RC MEDLINE=98012243; PubMed=9348339;
RA Comu S., Wang W., Olinak S., Ishwad P., Mi Z., Hempel J., Watkins S.,
RA Lagenaur C.F., Narayanan V.;
RT "The murine P84 neural adhesion molecule is SHPS-1, a member of the
RT phosphatase-binding protein family.";
RL J. Neurosci. 17:8702-8710(1997).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365,
RP N-GLYCOSYLATION, PHOSPHORYLATION ON TYROSINE RESIDUES, INTERACTION
RP WITH PTN6, AND TISSUE SPECIFICITY.
RC STRAIN=C57BL/6; TISSUE=Fetal thymus;
RX MEDLINE=98380500; PubMed=9712903;
RA Vellente A., Thibaudan E., Latour S.;
RT "High expression of inhibitory receptor SHPS-1 and its association
RT with protein tyrosine phosphatase SHP-1 in macrophages.";
RL J. Biol. Chem. 273:22719-22728(1998).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365, AND
RP N-GLYCOSYLATION.
RC STRAIN=129/SvJ; and C57BL/6; TISSUE=Brain, and Liver;
RX MEDLINE=20053880; PubMed=10585853;
RA Sano S.-I., Ohnishi H., Kubota M.;
RT "Gene structure of mouse Birt/SHS-1.";
RL Biochem. J. 344:667-675(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365.
RC STRAIN=ILS, and ISS;
RX MEDLINE=21363810; PubMed=11471062;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RT "High-throughput sequence identification of gene coding variants
RT within alcohol-related OTRs.";
RL Mamm. Genome 12:657-663(2001).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT PRO-194.
RC TISSUE=Placenta;
RA Wang H., Chen Z., Ullrich A.;
RT "Epidermal growth factor-induced association of SHP2 with mouse SHP-
RT alpha1.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [8]
RP FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=90152134; PubMed=2303162;
RA Chuang W., Lagenaur C.F.;
RT "Central nervous system antigen P84 can serve as a substrate for
RT neurite outgrowth.";
RL Dev. Biol. 137:219-232(1990).
RN [9]
RP N-GLYCOSYLATION, PHOSPHORYLATION BY JAK2 IN RESPONSE TO GROWTH
RP HORMONE, AND INTERACTIONS WITH JAK2 AND PTN1.
RX MEDLINE=98175985; PubMed=9507023;
RA Stoege M.R., Wang H., Ullrich A., Carter-Su C.;
RT "Growth hormone regulation of SHP and SHP-2 tyrosyl phosphorylation
RT and association.";
RL J. Biol. Chem. 273:7112-7117(1998).
RN [10]
RP INTERACTION WITH CD47, AND TISSUE SPECIFICITY.
RX MEDLINE=99091586; PubMed=9872987;
RA Jiang P., Lagenaur C.F., Narayanan V.;
RT "Integrin-associated protein is a ligand for the P84 neural adhesion
RT molecule.";
RL J. Biol. Chem. 274:559-562(1999).

```

```

RN [11]
RP INTERACTION WITH FGR.
RX MEDLINE=20130295; PubMed=10662797;
RA Gresham H.D., Dale B.M., Potter J.W., Chang P.W., Vines C.M.,
RA Lowell C.A., Lagenaur C.F., Willman C.L.;
RT "Negative regulation of phagocytosis in murine macrophages by the Src
RT kinase family member, Fgr.";
RL J. Exp. Med. 191:515-528(2000).
RN [12]
RP FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts
RP as docking protein and induces translocation of PTN6,
RP PTN1 and other binding partners from the cytosol to the
RP plasma membrane. Supports adhesion of cerebellar neurons, neurite
RP outgrowth and glial cell attachment. May play a key role in
RP intracellular signaling during synaptogenesis and in synaptic
RP function. Involved in the negative regulation of receptor tyrosine
RP kinase-coupled cellular responses induced by cell adhesion, growth
RP factors or insulin. Mediates negative regulation of phagocytosis,
RP mast cell activation and dendritic cell activation. CD47 binding
RP prevents maturation of immature dendritic cells and inhibits
RP cytokine production by mature dendritic cells (By similarity).
RN [13]
RP SUBUNIT: Binds PTN1 when tyrosine-phosphorylated, except in
RP macrophages, where it primarily binds PTN6. Binds GRB2 in vitro.
RP Binds FGR. Binds JAK2 irrespective of its phosphorylation status
RP and forms a stable complex. Binds SCAP1 and/or SCAP2. The
RP resulting complex recruits Fyb. Binds PRX2B (by similarity).
RN [14]
RP ALTERNATIVE PRODUCTS: Type I membrane protein.
RP Event=Alternative splicing; Named isoforms=3;
RP Comment=Additional isoforms seem to exist;
RP Name=1; Synonyms=a;
RP IsoId=p97797-1; Sequence=Displayed;
RP Name=2; Synonyms=a', Large; VSP_007032;
RP IsoId=p97797-2; Sequence=VSP_007031;
RP Name=3; Synonyms=b, Small;
RP IsoId=p97797-3; Sequence=VSP_007031;
RN [15]
RP TISSUE SPECIFICITY: Highly expressed in cerebral cortex, brain,
RP spinal cord, cerebellum and spleen, and at much lower levels in
RP kidney, thymus, heart, lung and liver. Within the cerebellum,
RP glomeruli in the granule cell layer. Detected in neurons of the
RP hippocampus and dentate gyrus, and in olfactory bulb. Not detected
RP in Purkinje cells. Highly expressed in the plexiform layers, optic
RP fiber layer and the outer segments of the photoreceptor layer in
RP the retina. Highly expressed in macrophages. Isoform 3 is detected
RP at very low levels in all tissues tested.
RN [16]
RP DEVELOPMENTAL STAGE: Highly expressed in the CNS of embryos from
RP day 7 to 17.
RN [17]
RP PTM: N-glycosylated.
RN [18]
RP PTM: Phosphorylated on tyrosine residues.
RN [19]
RP SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
RN [20]
RP SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
RN [21]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (See http://www.isb-sib.ch/announce/
RP or send an email to license@isb-sib.ch).
RN [22]
RP EMBL, D87967; BAA13520.1; -
RP EMBL, D87968; BAA13521.1; -
RP EMBL, D87969; BAA13522.1; -
RP EMBL, D87970; BAA13523.1; -
RP EMBL, D87971; BAA13524.1; -
RP EMBL, AF072543; AAC24886.1; -
RP EMBL, AF072544; AAC24887.1; -
RP EMBL, AF072545; AAC24888.1; -
RP EMBL, AB024500; BAA89290.1; JOINED.
RP EMBL, AB024501; BAA89290.1; JOINED.
RP EMBL, AB024502; BAA89290.1; JOINED.
RP EMBL, AB024503; BAA89290.1; JOINED.
RP EMBL, AB024504; BAA89290.1; JOINED.
RP EMBL, AB024505; BAA89290.1; JOINED.

```



DR EMBL; AB024506; BAA69290.1; JOINED.  
 DR EMBL; AB018194; BAA76555.1; -.  
 DR EMBL; AB024507; BAA69289.1; -.  
 DR EMBL; AB024500; BAA69289.1; JOINED.  
 DR EMBL; AB024501; BAA69289.1; JOINED.  
 DR EMBL; AB024502; BAA69289.1; JOINED.  
 DR EMBL; AB024503; BAA69289.1; JOINED.  
 DR EMBL; AB024504; BAA69289.1; JOINED.  
 DR EMBL; AB024505; BAA69289.1; JOINED.  
 DR EMBL; AB024506; BAA69289.1; JOINED.  
 DR EMBL; AF332079; AAK56107.1; -.  
 DR EMBL; AF332080; AAK56108.1; -.  
 DR EMBL; Y10349; CAA71375.1; -.  
 DR PIR; JCS289; JCS289.  
 DR MGD; MGI:108563; Ptpn1.  
 DR GO; GO:0006580; F:cytoskeletal regulator activity; IMP.  
 DR GO; GO:0045309; F:phosphoprotein amino acid binding; IPT.  
 DR GO; GO:0005515; F:protein binding; IPT.  
 DR GO; GO:0007015; P:actin filament organization; IMP.  
 DR GO; GO:0006928; P:cell motility; IMP.  
 DR GO; GO:0007160; P:cell-matrix adhesion; IMP.

Query Match 6.0%; Score 203.5; DB 1; Length 513;  
 Best Local Similarity 22.3%; Pred. No. 7.9e-06;  
 Matches 92; Conservative 62; Mismatches 156; Indels 103; Gaps 18;

OY 248 SKSWITFDLKNKVEVVKAVTODPKQWKGKLPMLHLTPQALP-----QYAGSGNULTALE 302  
 DB 21 SASCECTCVTKELKLVTPKESVVAAGDSTVLNCTLSLPVGPPIKRYRGVGGSRLLIY 80  
 OY 303 AKTGKLGQEVNLVWMRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKREKRVWYLN 362  
 DB 81 SFTGHPFRVTVN---SDATKRN-----NMDFSIHISN-----VTP 113  
 OY 363 PEAGMQLDSDGCVLLESNIKVLPTWSTVEPESCKTKTCCPCAPPELL--GGRSVF 420  
 DB 114 EDAGTYVVKFKQK-----PSEEDPT-----ETQSGGGEVY 144  
 OY 421 LF---PK---PKDLMISRPEVTVVNVSH--EDPEVKFMYVGVGVHNAKTKPR 471  
 DB 145 VLAKSPPEVSGPADRGIPDQKVNFTC---KSHQFSRNITLTKFKDGGSLHLETTVN 200  
 OY 472 EEOYNSTYRVVSVLTVLHQDWLNGEKYKVNKPAPIEKITSKAGQPREBOVYTLR 531  
 DB 201 PSKKNVSYNISSTYRV---LNSMDVSKVYICEVAHITLDR--SPLKGINLNSFTKVS 254  
 OY 532 PS-----RDELTKQVSLTCLVKGYPDSIDIAWESNQPENNYKTPPV--DSDSGF 584  
 DB 255 PTVKVTQOSPMSQVNLTCRAERYPBDLQILWLENGVSRN--DTPKNTKNTDGTYN 312  
 OY 585 LYSKLTVDKSRMOQGNVSCSMHE--ALNHYT-----QKSLSLSPG 625  
 DB 313 YTSFLVNVSSAHRVDVFTCVQKHQDQPAITRNHTVGLAASSDQSGMQTTPG 365

RESULT 59  
 SHS1\_BOVIN STANDARD; PRT; 506 AA.  
 AC 046631; 046632;

DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Protein-tyrosine phosphatase non-receptor type substrate 1 precursor  
 (SHP substrate-1) (SHP-1) (Inhibitory receptor SHP-1) (Signal-  
 DE regulatory protein alpha-1) (SIRP-alpha-1) (MyD-1 antigen).  
 GN PTNS1 OR SHP1 OR SIRP OR MYD1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANTS SER-23; ALA-28; LEU-61; ARG-70;

RP HIS-120; 125-GLN; 127-GLY; 129-HIS; 132-VAL; ASN-145; VAL-153;  
 RP ASP-203; ARG-261; LEU-302; LEU-316; ARG-337; ASN-367; LEU-422; PHE-429  
 RP AND GLU-433.  
 RC STRAIN=Friestian; TISSUE=Peripheral blood;  
 RA BROOKE G.P., PARSONS K.R., HOWARD C.J.;  
 RX MEDLINE=98143722; PubMed=9485180;  
 RT "Cloning of two members of the SIRP alpha family of protein tyrosine  
 RT phosphatase binding proteins in cattle that are expressed on monocytes  
 RT and a subpopulation of dendritic cells and which mediate binding to  
 RT CD4 T cells.";  
 RL Eur. J. Immunol. 28:1-11(1998).  
 CC -1- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts  
 CC as docking protein and induces translocation of PTPN6,  
 CC PTPN11 and other binding partners from the cytosol to the  
 CC plasma membrane. Supports adhesion of cerebellar neurons, neurite  
 CC outgrowth and glial cell attachment. May play a key role in  
 CC intracellular signaling during synaptogenesis and in synaptic  
 CC function. Involved in the negative regulation of  
 CC receptor tyrosine kinase-coupled cellular responses induced by  
 CC cell adhesion, growth factors or insulin. Mediates negative  
 CC regulation of phagocytosis, mast cell activation and dendritic  
 CC cell activation. CD47 binding prevents maturation of immature  
 CC dendritic cells and inhibits cytokine production by mature  
 CC dendritic cells (By similarity).  
 CC -1- SUBUNIT: Binds PTPN11 when tyrosine-phosphorylated, except in  
 CC Bnde JNK2 irrespective of its phosphorylation status and forms a  
 CC scabie complex. Binds SCAP1 and/or SCAP2. The resulting complex  
 CC recruits Fyb. Binds Fcr and PTK2B (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in spleen macrophages.  
 CC -1- DETECTED IN skin dendritic cells.  
 CC -1- PTM: Phosphorylated on tyrosine residues (By similarity).  
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by, and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
 DR EMBL; Y11045; CAA71942.1; -.  
 DR EMBL; Y11046; CAA71943.1; -.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003597; Ig-cl.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00047; Ig; 3.  
 DR SMART; SM00407; IgC1; 2.  
 DR PROSITE; PSS0835; IG\_LIKE; 3.  
 DR PROSITE; PSS0290; IG\_MHC; FALSE NEG.  
 KW Repeat; Signal; Transmembrane; Immunoglobulin domain; SH3-binding;  
 KW Glycoprotein; Phosphorylation; Polymorphism.  
 FT SIGNAL 1 29  
 FT CHAIN 30 506  
 FT DOMAIN 30 371  
 FT TRANSLEM 372 392  
 FT DOMAIN 393 506  
 FT DOMAIN 30 145  
 FT DOMAIN 148 248  
 FT DOMAIN 255 348  
 FT DISULFID 121 121  
 FT DISULFID 170 228  
 FT DISULFID 273 331  
 FT SITE 432 435  
 FT SITE 441 446  
 FT SITE 455 458  
 FT SITE 472 475  
 FT SITE 498 501  
 FT MOD\_RES 431 431  
 TYPE SUBSTRATE 1.  
 EXTRACELLULAR (POTENTIAL).  
 POTENTIAL.  
 CYTOPLASMIC (POTENTIAL).  
 IG-LIKE V-TYPE.  
 IG-LIKE C1-TYPE 1.  
 IG-LIKE C1-TYPE 2.  
 POTENTIAL.  
 POTENTIAL.  
 SH2-BINDING (POTENTIAL).  
 SH3-BINDING (POTENTIAL).  
 SH2-BINDING (POTENTIAL).  
 SH2-BINDING (POTENTIAL).  
 SH2-BINDING (POTENTIAL).  
 PHOSPHORYLATION (BY TYR-KINASES)



	FT	MOD_RES	455	(POTENTIAL) PHOSPHORYLATION (BY TYR-KINASES)
FT	MOD_RES	472	472	(POTENTIAL) PHOSPHORYLATION (BY TYR-KINASES) (BY SIMILARITY).
FT	MOD_RES	498	498	PHOSPHORYLATION (BY TYR-KINASES) (BY SIMILARITY).
FT	CARBONYD	92	92	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	167	167	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	179	179	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	204	204	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	210	210	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	246	246	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	270	270	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	292	292	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	311	311	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	319	319	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	344	344	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	365	365	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD	368	368	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	VARIANT	23	23	N -> S.
FT	VARIANT	28	28	T -> A.
FT	VARIANT	61	61	S -> L.
FT	VARIANT	70	70	K -> R.
FT	VARIANT	120	120	Y -> H.
FT	VARIANT	125	125	R -> Q.
FT	VARIANT	127	127	E -> G.
FT	VARIANT	129	129	M -> H.
FT	VARIANT	132	132	R -> V.
FT	VARIANT	145	145	S -> N.
FT	VARIANT	153	153	L -> V.
FT	VARIANT	203	203	N -> D.
FT	VARIANT	261	261	G -> F.
FT	VARIANT	302	302	F -> L.
FT	VARIANT	316	316	F -> R.
FT	VARIANT	337	337	G -> L.
FT	VARIANT	367	367	S -> N.
FT	VARIANT	422	422	Q -> L.
FT	VARIANT	429	429	I -> F.
FT	VARIANT	433	433	D -> E.
SEQ	SEQUENCE	506 AA;	55093 MW;	6B7EJ10677FCF9CB CRC64;
Query Match	Beet Local Similarity	5.5%;	Score 187.5;	DB 1; Length 506;
Matches	89;	Conservative	70;	Mismatches 168; Indels 99; Gaps 17;
Oy	246	SSSKSMIFDELKDKNKVSVKRVYQDPKLQMGGKLPHHLPLPOLP-----QYAGSGNLTLLA	300	
Db	20	AASNAM-IGTADGDELOVIQPRSRVSVAAGETATLICTVTSLSPVGPIKMFKGTV-----	73	
Oy	301	LEAKTGK-LHOENVLVVMRATOL-----OKNLTCVWGPTSPKMLSLLENKEAKVSKR	354	
Db	74	----PGREFIYSQKAPFPFRVYNVDARKN-----NMDFSIRISN-----	110	
Oy	355	EKPVVVLNPEAGMOCCLSDSQVLLLESNIKVLPTWSTPVEBKSCDKTH-TCPPCPAEL	413	
Db	111	-----ITPADAGVYYC-----VKFKKERGDGMFEKSGPGTHLTVSAKRSPPV	152	
Oy	414	LGPSVFLPPRPKKTLMISRTPETCVVNVDSH--EDBEVKFNMYVDGVENNAKTXPR	471	
Db	153	LSGPV-----RAIPDETIVFTCTSHGPSRPNISLKMFKNGNELSASQTSVD	199	
Oy	472	EEQVSTYRVVSVLTLVHQDWMLNGEVKYCKVSNKALPA-PLEKTIISAKAQGPREPQYT	529	
Db	200	PEDNNVVSINSITTYVLAATGTGVSQCIVEAHVTLQGSPV-----RGIANISEIIR	252	
Oy	530	LPSPHDEL-----KNQVSLTCLVKGFYPSPDIADVEMESNGOPENNKYTTTPYLDSGSFF	584	
Db	253	VPPTL-EITGSPSPASGNQVNWVTCQWKFPRIHLQTLWELEGNMSRTREASVFVENKDGTFN	311	
Oy	585	LYSKLTYDVKSRRKQQGNVSSCSYMHALNHNHYQXLSLSPQLQDLDECAEAODEL----	640	
Db	312	QTSWELVNSSAHRAELAVVLTICQVHHNG-----QPRLVSKNHLEVASAPQKDDOTQTPBN	365	

ID	DB	DSNMTS	641	DGLMTT	646
CV			641	DGLMTT	646
DB		366	DSNMTS	371	
RESULT	60				
VCAM1	RAT				
ID	VCAM1	RAT	STANDARD;	PRT;	739 AA.
AC	P29534;				
DT	01-APR-1993	(Rel. 25, Created)			
DT	01-APR-1993	(Rel. 25, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	Vascular cell adhesion protein 1 precursor (V-CAM 1).				
GN	VCAM1 OR VCAM-1.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxId=10116;				
RY	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Lung;				
RX	MEDLINE=92181437; PubMed=1371918;				
RA	Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wysk M.,				
RA	Burkly L., Myake K., Kincaid P., Lobb R.;				
RT	"Cloning of murine and rat vascular cell adhesion molecule-1.";				
RL	Biochem Biophys. Res. Commun. 183:163-169(1992).				
CC	-1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION				
CC	IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1				
CC	INTEGRIN VL44 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL				
CC	TRANSDUCTION. THE VCAM1/VL44 INTERACTION MAY PLAY A				
CC	PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE				
CC	EMIGRATION TO SITES OF INFLAMMATION.				
CC	-1- SUBCELLULAR LOCATION: Type 1 membrane protein.				
CC	-1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as				
CC	well as on macrophage-like and dendritic cell types in both normal				
CC	and inflamed tissue.				
CC	-1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/or send an email to license@isb-sib.ch">http://www.isb.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; M84488; AAA42332.1; -.				
DR	PIR; J50675; J50675.				
DR	HSSP; P19930.1; VCAM.				
DR	InterPro; IPR0033987; ICAM VCAM-1.				
DR	InterPro; IPR007110; Ig-like.				
DR	InterPro; IPR003598; Ig_c2.				
DR	InterPro; IPR003989; VCAM-1.				
DR	Pfam; PF00047; Ig; 5.				
DR	PRINTS; PRO1472; ICAMVCAM1.				
DR	PRINTS; PRO1474; VCAM1.				
DR	SMART; SM00408; IGC2; 3.				
DR	PROSITE; PS50835; IG_LIKE; 5.				
KW	Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;				
RV	Repeat; Signal.				
FT	SIGNAL	1	24		
FT	CHAIN	25	739		
FT	DOMAIN	25	698		
FT	TRANSMEM	699	720		
FT	DOMAIN	721	739		
FT	DOMAIN	25	111		
FT	DOMAIN	119	212		
FT	DOMAIN	223	309		
FT	DOMAIN	312	3		



```

FT DISULFID 47 95 BY SIMILARITY.
FT DISULFID 52 99 BY SIMILARITY.
FT DISULFID 137 195 BY SIMILARITY.
FT CARBOHYD 273 273 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 561 561 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 650 650 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 739 AA; 81246 MW; 566855A1A1B100C CRC64;

Query Match 5.3%; Score 182.5; DB 1; Length 739;
Base Local Similarity 20.8%; Pred.No.0.00028;
Matches 124; Conservative 87; Mismatches 195; Indels 191; Gaps 28;

OY 33 KGDYELTCTASOKSIOFHWKNSNQIKILGNGSFLFKGPKSLNDRASRRSLMDQNF 92
DB EGAANTWTCASGCLPAPRIFMSK-----LNGVLQQL-----SNA 274
OY 93 PLIINKLIEDSDTYICE---VEDQKEVOLLV-----FGLTANSPTLLQGSITLT 142
DB 275 TLTLIAMEDESGIYCEGVNLVGRDKTEVELIVQEKFPYVDISPGQVAAQVGDVYLT 334
OY 143 LESPPGSSPSVOCRRPRGNKIOG-----GKTLVSQLELDQDSTWTCTVLQNKRYEF 195
DB 335 CAAVGCDSPSPSWRTQUTDSPINGEVRDGAFTSLTLSPVGEVDEHSYLCYTCORRKLK 394
OY 196 KIDIVLVLFQKASIVYKKEGQVEFSPFLAFTVYKLTGSGELMWQERASSKSWITFD 255
DB 395 TIQVEVYF-----PEDEITISGPLVH----- 417
OY 256 LKNEVSVKRVYTDPKLOMKKPLHLTLPLQALPQVAGSNLTLALAKTKLHQEVNLV 315
DB 418 -----GRPVYVNCVPMVYFPD-----HLEIELL 441
OY 316 VMRATQLOKULTCEVWGPTSPKMLSLKLENKEAVSREKRVWVLEAGMWCILSDS 375
DB 442 KGETTLNKKFIREI-GTKS-----LETKSLMT-----FIPTA-----EDT 477
OY 376 GOVLLESNIKVLPTWSTVEPEKSCDKTATCPCPAPRLLGSGSVLPF-PKPKDTIMISR 434
DB 478 GKALV--CLAKLHSSQMESEKQROSTQTLVYVNAFK---EPTIWSSPVPREBSPN- 531
OY 435 TPVATCVVDVSHEDPEVKFMYVGVGVHNAKTKPREQYNSTYRVVSVLTLVHQDWLN 494
DB 532 ---LTC---SSDGPPTKILM---SRQLKNGELOPLSQ---NTL---LSFMATKMED--- 573
OY 495 GKREYCKXSNKALPRPIEKITSK-----ARGQPREPVYTLPPSRDELTKQVSLTCLV 548
DB 574 SGIVYCEGINEA-----GISKSKVELLIQSSSKDIDL-TVPBSKYVEGDTVIISCTC 625
OY 549 KGFPYSDIAVEMESNGQPPENNYKTPPYLDS-DGSFFLYSKLTVDKSRWQGNFSC 604
DB 626 -GSVPEIIT-----LKKXAKTODMWLKSVNGSY-----TIKQALQDAGYIEC 668

RESULT 61
PGBM_HUMAN STANDARD; PRT; 4391 AA.
AC P98150; 016827; 09H3V5.
O1-OCT-1996 (Rel. 34, Created)
D1 28-FEB-2003 (Rel. 41, Last sequence update)
D7 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112994; PubMed=1730768;
Kallunki P., Tryggvason K.;

```

```

RT RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor.";
RT J. Cell Biol. 116:559-571(1992).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Colon, and Skin;
RC MEDLINE=92235084; PubMed=1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor.";
RT J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RN SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SISI TYR-1532.
RX MEDLINE=2053141; PubMed=1101850;
RA Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA Beignon P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA Samson D., Urtubera J.A., Lehmann-Horn F., Weissenbach J.,
RA Hentati F., Fontaine B.;
RT "Perlecan, the major proteoglycan of basement membranes, is altered in
RT patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia)."
RT Nat. Genet. 26:480-483(2000).
RN [4]
RN SEQUENCE OF 1016-1470 FROM N.A.
RP TISSUE=Colon;
RC MEDLINE=91365376; PubMed=1679749;
RA Dodge G.R., Kovalsky I., Chu M.L., Hassell J.R., McBride O.W.,
RA Yi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RT Genomics 10:673-680(1991).
RN [5]
RN SEQUENCE OF 890-1396 FROM N.A.
RP TISSUE=Fibroblasts;
RC MEDLINE=92120660; PubMed=1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Keestila M., Shows T.B.,
RA Tryggvason K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to 1p36.1--p35 and identification of
RT a BamHI restriction fragment length polymorphism.";
RT Genomics 11:389-396(1991).
RN [6]
RN SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=94052171; PubMed=8234307;
RA Cohen I.R., Graessle S., Murdoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter.";
RT Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN [7]
RN CARBOHYDRATE-LINKAGE SITE ASN-2121.
RP MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT "Identification and quantification of N-linked glycoproteins using
RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RT Nat. Biotechnol. 21:660-666(2003).
RN [8]
RN FUNCTION: This protein is an integral component of basement
RN membrane. It is responsible for the fixed negative electrostatic
RN charge and is involved in the charge-selective ultrafiltration
RN properties. It serves as an attachment substrate for cells.
RN -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
RN dimers or stellate structures. It interacts with other basement
RN membrane components such as laminin, prolargin and collagen type
RN IV.
RN -1- SUBCELLULAR LOCATION: Extracellular.
RN -1- TISSUE SPECIFICITY: Found in the basement membranes.
RN -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
RN AND O-LINKED OLIGOSACCHARIDES.
RN -1- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
RN syndrome (SUS1) [MIM:255800]; a rare autosomal recessive disorder
CC

```



```

CC      characterized by permanent myotonia (prolonged failure of muscle
CC      relaxation) and skeletal dysplasia, resulting in reduced stature,
CC      kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC      -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC      -1- SIMILARITY: Contains 11 laminin EGF-like domains.
CC      -1- SIMILARITY: Contains 3 laminin IV domains.
CC      -1- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC      -1- SIMILARITY: Contains 3 laminin G-like domains.
CC      -1- SIMILARITY: Contains 4 EGF-like domains.
CC      -1- SIMILARITY: Contains 1 SEA domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X62515; CAA44373.1; -
DR      EMBL; M85289; AA52700.1; -
DR      EMBL; AL445795; CAC18534.1; -
DR      EMBL; W64283; AA52693.1; -
DR      EMBL; S76436; AAB21121.2; -
DR      EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR      PIR; A38096; A38096.
DR      HSP; P00740; 1EDM.
DR      Stena-2DPAGE; P98160; -
DR      Genew; HGNC:5273; HSPG2.
DR      MIM; 142461; -
DR      MIM; 255800; -
DR      InterPro; IPR008985; ConA_like_1ec_g1.
DR      InterPro; IPR000742; EGF_3.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR006210; IEGF.
DR      InterPro; IPR007110; IG-like.
DR      InterPro; IPR003599; IG.
DR      InterPro; IPR003598; IG_C2.
DR      InterPro; IPR003596; IG_Y.
DR      InterPro; IPR000034; Laminin_B.
DR      InterPro; IPR002049; Laminin_EGF.
DR      InterPro; IPR001791; Laminin_G.
DR      InterPro; IPR002172; LDL_receptor_A.
DR      InterPro; IPR000082; SEA_domain.
DR      Pfam; PR00008; EGF_4.
DR      Pfam; PR00047; IG_22.
DR      Pfam; PR00052; Laminin_B_3.
DR      Pfam; PR00053; Laminin_EGF_7.
DR      Pfam; PR00054; Laminin_G_3.
DR      Pfam; PR00057; ldl_recept_a; 4.
DR      Pfam; PR01390; SEA_1.
DR      PRINTS; PR00261; LDLRECEPTOR.
DR      ProDom; PD003031; Laminin_B_3.
DR      SMART; SM00181; EGF_15.
DR      SMART; SM00180; EGF_Lam; 12.
DR      SMART; SM00409; IG_22.
DR      SMART; SM00408; IEG2; 21.
DR      SMART; SM00406; IGV_7.
DR      SMART; SM00281; Lamb; 3.
DR      SMART; SM00282; LamG; 3.
DR      SMART; SM00192; LDLa; 4.
DR      SMART; SM00200; SEA_1.
DR      PROSITE; PS00022; EGF_1; 9.
DR      PROSITE; PS01186; EGF_2; 6.
DR      PROSITE; PS00026; EGF_3; 4.
DR      PROSITE; PS00835; IG_Like; 22.
DR      PROSITE; PS00025; IAM_G_DOMAIN; 3.
DR      PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR      PROSITE; PS01209; LDLa_1; 4.
DR      PROSITE; PS00068; LDLa_2; 4.
DR      PROSITE; PS00024; SEA_1.
KW      Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW      Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;

```

```

KW      Extracellular matrix; EGF-like domain; Disease mutation.
FT      SIGNAL 1 21
FT      CHAIN 22 4391
FT      DOMAIN 80 194
FT      DOMAIN 198 235
FT      DOMAIN 284 360
FT      DOMAIN 324 320
FT      DOMAIN 367 404
FT      DOMAIN 405 504
FT      DOMAIN 521 530
FT      DOMAIN 531 730
FT      DOMAIN 731 763
FT      DOMAIN 764 813
FT      DOMAIN 814 871
FT      DOMAIN 879 923
FT      DOMAIN 924 933
FT      DOMAIN 934 1125
FT      DOMAIN 1126 1158
FT      DOMAIN 1159 1208
FT      DOMAIN 1209 1265
FT      DOMAIN 1275 1324
FT      DOMAIN 1325 1334
FT      DOMAIN 1335 1529
FT      DOMAIN 1530 1562
FT      DOMAIN 1563 1612
FT      DOMAIN 1613 1771
FT      DOMAIN 1677 1771
FT      DOMAIN 1772 1865
FT      DOMAIN 1866 1955
FT      DOMAIN 1956 2051
FT      DOMAIN 2052 2151
FT      DOMAIN 2152 2244
FT      DOMAIN 2245 2340
FT      DOMAIN 2341 2436
FT      DOMAIN 2437 2533
FT      DOMAIN 2534 2629
FT      DOMAIN 2630 2726
FT      DOMAIN 2630 2726
Query Match 5.2%; Score 178; DB 1; Length 4391;
Best Local Similarity 20.7%; Pred. No. 0.0058;
Matches 158; Conservative 97; Mismatches 292; Indels 218; Gaps 40;
QY      5 VPRRLVLLVQLALLPA--ATQGNV--VLGKGP-----TVLLTCTASQKSIQTHW 53
DB      2477 LPARQVHGSRRLRLQVTPADSGEYVCRVVGSGTQESVLTVTIQRLSGSHSGVAV-- 2534
QY      54 KSNQIKILNGSGFLTKGPSKLANDRADSRSS--LMDQ--GNFP-----LITKUL 99
DB      2535 ----PRLSSASLANGHTLDLNCVASQAPHTTWYKRGSGLPSRQIYGSRLRIFQV 2590
QY      100 KIEDSDTYICEYED--QKEEVQLVFGLTANSDTHLQGO--SLTTLTSSP---GSSPS 152
DB      2591 TPADSGEYVCHVSNAGSRETSLIV-----TIGSSSHVSVSPPIRISSSPT 2640
QY      153 V-----QC---RSP-----RGKNIQG---GKITSVSLQLSDGSTWTCTYVLQ 188
DB      2641 VVEGOTLDLNCVVAQPOAIIITWYKRGSGLPSRHQTHSHRLHQMVAADSGEYVCRRANN 2700
QY      189 NOKVVEFKIDIV-----LAFQKASIVYKKEGVOVFFSPPLAFVTEKLGVS 235
DB      2701 NIDLEASIVISVSASGSPAPGSMPIRISSSHVABEGTLDLN-----CVPRQAH 2755
QY      236 GELMWQAEARASSKSWITFDLKNKEVSVKRVYTOPKIQMGKKLPLHLTLPOALPOYAGSG 295
DB      2756 AQVTHKRGKGLPSHHQT-----RGRRLRLRHVHSP-----ADSG 2789
QY      236 NLTLLAEKTKGLHGVNVLVYMKRATQLOK-----NITC 328
DB      2790 EYVCRVWSSGGLFASV-LVTIEASGSAVHPAPGAPPIRIEBSRVAEGOTLDLKC 2848
QY      329 EVMGPTSPKLMSTLKEKKAIVSKREKPVWVLN-----PEAGMOCCLSDSGQVLTESNI 384

```



```

Db      2849  VPPGAAHQVTHKRGKGLPAR-HQVHGRPLRLNVSADSGEYSCQVTSST-LEASV 2906
Oy      385  KVLPTWSTPVEPKSCDKHTTCP-PCPAPELLGGPVEFLPPPKKDTLMISRTPEVTCVV 443
Db      2907  LV-----TIPESS-----PGPIPA-----GLAQPIYEASSSHVTEGOTDLNLCVVP 2949
Oy      444  DVSHEDPEVKFNVDVGVENHNAKTRPEEQYNSTYRVSVLTVLHQMNGKEYCKVSV 503
Db      2950  GOAH-----AQVTHWYKRG-----GSLPARHQTHGSLRLHLVSPA-----DSGEYVCRPA 2994
Oy      504  NKALP-----APIEKTISKAKGP-----REPOVTLPPSRBELTNQVSLTCLV-KGFPYSDI 556
Db      2995  SGPGGEQKSTFTVTVPPSGSSRYLRSPVISIDPPSSITVQCGQDASFKCLIHDAAP--I 3052
Oy      557  AVEMESNGQP-ENNYKTPPVLDSDGSFFLYSKLTVDKSRMQQGVFSC-----S 605
Db      3053  SLEWTKRQDELDNVHISP-----NGSI-----ITVGTREPNHGTTCVASNAYGVAQS 3102
Oy      606  VMHEALHNHYTKSLTSLSP-----GLQDETCAEQDDELGLWT 645
Db      3103  VVNLISVHGPPTVSVLPBGPVWVKGVKAVTLKCVSAGSEPRSSARWT 3147

```

## RESULT 62

LAC1\_MOUSE STANDARD; PRT; 105 AA.

```

AC      P01843;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Ig lambda-1 chain C region.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxId=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=83014953; PubMed=6812053;
RA      Seising E., Miller J., Wilson R., Storb U.;
RT      "Evolution of mouse immunoglobulin lambda genes.";
RL      Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).
RN      [2]
RP      SEQUENCE FROM N.A. (MOPC 315).
RX      MEDLINE=81148606; PubMed=6259534;
RA      Botwell A.L.M., Paekind M., Schwartz R.C., Sonenshein G.E.,
RA      Gefter M.L., Baltimore D.;
RT      "Dual expression of lambda genes in the MOPC-315 plasmacytoma.";
RL      Nature 290:65-67(1981).
RN      [3]
RP      SEQUENCE FROM N.A. (S43).
RX      MEDLINE=82220143; PubMed=6283385;
RA      Botwell A.L.M., Paekind M., Reth M., Imanishi-Kari T., Rajewsky K.,
RA      Baltimore D.;
RT      "Somatic variants of murine immunoglobulin lambda light chains.";
RL      Nature 298:380-382(1982).
RN      [4]
RP      SEQUENCE (MYELOMA PROTEIN MOPC 104E).
RX      MEDLINE=71107854; PubMed=5276767;
RA      Appella E.;
RT      "Amino acid sequences of two mouse immunoglobulin lambda chains.";
RL      Proc. Natl. Acad. Sci. U.S.A. 68:590-594(1971).
CC      -1- MISCELLANEOUS: The MOPC 315 cell line produces 2 light chains, 1
CC      normal lambda-2 chain and 1 abnormal lambda-1 chain that is
CC      missing a large part of the V region. The C region sequence (shown
CC      here) appears completely normal.
CC      -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

```

CC      EMBL; J00582; AAAS1636.1; -.
CC      EMBL; J00587; AAB59672.1; -.
CC      PIR; A93922; LIMS.
CC      PDB; 1JNH; 06-FEB-02.
CC      InterPro; IPR007110; Ig-like.
CC      InterPro; IPR003597; Ig_c1.
CC      InterPro; IPR003006; Ig_MHC.
CC      Pfam; PF00047; Ig; 1.
CC      SMART; SM00407; Igcl; 1.
CC      PROSITE; PS50835; IG_LIKE; 1.
CC      PROSITE; PS00290; IG_MHC; 1.
CC      KX      Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
CC      FT      NON_TER      1      1
CC      FT      DOMAIN      1      1
CC      FT      DISULFID      27      86      IG-LIKE.
CC      FT      DISULFID      104      104
CC      FT      CONFLICT      19      20      INTERCHAIN (WITH HEAVY CHAIN).
CC      FT      CONFLICT      56      56      ET -> TE (IN REF. 4).
CC      FT      CONFLICT      75      75      Q -> E (IN REF. 4).
CC      FT      CONFLICT      81      82      MISSING (IN REF. 4).
CC      FT      CONFLICT      85      85      HS -> SH (IN REF. 4).
CC      FT      CONFLICT      96      96      S -> SS (IN REF. 4).
CC      FT      CONFLICT      96      96      E -> Q (IN REF. 4).
CC      SQ      SEQUENCE      105 AA; 11575 MW; A89F2B09BCFCA018 CRC64;

```

Query Match 5.2%; Score 177.5; DB 1; Length 105;  
 Best Local Similarity 35.9%; Pred. No. 4.6e-05;  
 Matches 37; Conservative 19; Mismatches 43; Indels 5; Gaps 3;

```

Oy      521  QPR-EPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNGQP-ENNYKTPPVLD 578
Db      1  QPKSSPVTLTFPPSESELETNKATLVCTITDFYGVGVTVMKVDGTFVQGMETPTPSKQ 60
Oy      579  SDGSFFLYSKLTVDKSRMQQGVFSCVMHEALHNHYTKSLS 621
Db      61  SNKRYMASSYTLTLPARAMERHSSYSCQVTHE--GHTVERKSL 100

```

## RESULT 63

UN89\_CABEL STANDARD; PRT; 6632 AA.

```

ID      UN89_CABEL
AC      001761; Q17362;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN      UNC-89 OR C09D1.1.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxId=6239;
RN      [1]
RP      SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC      STRAIN=Bristol N2;
RX      MEDLINE=96180278; PubMed=8603916;
RA      Beilan G.M., Tinley T.L., Tang X., Borodovsky M.;
RT      "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT      assembly, encodes a giant modular protein composed of Ig and signal
RT      transduction domains.";
RL      J. Cell Biol. 132:835-848(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96180278; PubMed=8603916;
RA      Beilan G.M., Tinley T.L., Tang X., Borodovsky M.;
RT      "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT      assembly, encodes a giant modular protein composed of Ig and signal
RT      transduction domains.";
RL      J. Cell Biol. 132:835-848(1996).
RN      [3]
RP      Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN      REVIEWS.
RA      Waterston R.;
RL      Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: Structural component of the muscle M-line. Myofibrilment
CC      lattice assembly begins with positional cues laid down in the
CC      basement membrane and muscle cell membrane. UNC-89 responds to

```



these signals, localizes, and then participates in assembling an M-line.

- TISSUE SPECIFICITY: Localizes to the middle of A-bands.

- SIMILARITY: Contains 1 DBL-homology (DH) domain.

- SIMILARITY: Contains 1 fibronectin type III domain.

- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.

- SIMILARITY: Contains 1 PH domain.

- SIMILARITY: Contains 5 RCHD domains.

- SIMILARITY: Contains 1 SH3 domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announce/> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

-----

EMBL; U3058; AAB00542.1; -

EMBL; AF003131; AAB54132.2; -

PDB; 1FHO; 20-DEC-00

WormPep; C09D1.1; CB30426.

InterPro; IPR008957; FN\_III-like.

InterPro; IPR003961; FN\_III.

InterPro; IPR007110; Ig-like.

InterPro; IPR003598; Ig\_C2.

InterPro; IPR003006; Ig\_MHC.

InterPro; IPR001849; PH.

InterPro; IPR007850; RCHD.

InterPro; IPR000219; RhGEF.

InterPro; IPR001452; SH3.

Pfam; PF000041; fn3; 1.

Pfam; PF000047; ig; 47.

Pfam; PF00169; ph; 1.

Pfam; PF05177; RCHD; 5.

Pfam; PF00621; RhGEF; 1.

Pfam; PF00018; SH3; 1.

SMART; SM00408; IGC2; 23.

SMART; SM00325; RhGEF; 1.

SMART; SM00326; SH3; 1.

PROSITE; PS50010; DH\_2; 1.

PROSITE; PS50835; Ig\_LIKE; 49.

PROSITE; PS50003; PH\_DOMAIN; 1.

PROSITE; PS50002; SH3; 1.

Muscle protein; Immunoglobulin domain; Repeat; SH3 domain; 3D-structure.

FT DOMAIN 63 127 SH3.

FT DOMAIN 152 330 DH.

FT DOMAIN 342 498 PH.

FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.

FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.

FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.

FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.

FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.

FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.

FT DOMAIN 1272 1315 THR-RICH.

FT DOMAIN 1375 1475 RCHD 1.

FT DOMAIN 1479 1685 RCHD 2.

FT DOMAIN 1597 1955 RCHD 3.

FT DOMAIN 1700 1799 RCHD 4.

FT DOMAIN 1800 1982 RCHD 5.

FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.

FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.

FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.

FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.

FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.

FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.

FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.

FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.

FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.

FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.

FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.

FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.

FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.

FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.

FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.

FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.

FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.

FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.

FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.

FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.

FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.

FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.

FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.

FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.

FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.

FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.

FT DOMAIN 4678 4771 IG-LIKE C2-TYPE 33.

FT DOMAIN 4861 4961 IG-LIKE C2-TYPE 34.

FT DOMAIN 4961 5057 IG-LIKE C2-TYPE 35.

FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 36.

FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 37.

FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 38.

FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 39.

FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 40.

FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 41.

FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 42.

FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 43.

FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 44.

FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 45.

FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 46.

FT DOMAIN 6275 6368 IG-LIKE C2-TYPE 47.

FT DOMAIN 6413 6502 FIBRONECTIN TYPE-III.

FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 48.

FT DISULFID 6507 6596 IG-LIKE C2-TYPE 49.

FT DISULFID 568 621 POTENTIAL.

FT DISULFID 2908 2975 POTENTIAL.

FT DISULFID 3015 3065 POTENTIAL.

FT DISULFID 3707 3759 POTENTIAL.

FT DISULFID 3826 3890 POTENTIAL.

FT DISULFID 5092 5157 POTENTIAL.

FT DISULFID 5288 5350 POTENTIAL.

FT DISULFID 5508 5560 POTENTIAL.

FT DISULFID 5616 5669 POTENTIAL.

FT DISULFID 5722 5764 POTENTIAL.

FT DISULFID 5836 5901 POTENTIAL.

FT DISULFID 5946 5998 POTENTIAL.

FT DISULFID 6036 6171 POTENTIAL.

FT DISULFID 6421 6486 POTENTIAL.

FT CONFLICT 2137 2137 A -> P (IN REF. 1).

FT CONFLICT 2245 2247 AKA -> PKP (IN REF. 1).

FT CONFLICT 2258 2258 A -> P (IN REF. 1).

FT CONFLICT 2284 2284 E -> G (IN REF. 1).

FT CONFLICT 2287 2287 M -> I (IN REF. 1).

FT CONFLICT 3531 3531 A -> G (IN REF. 1).

FT CONFLICT 3531 3531 A -> G (IN REF. 1).

FT CONFLICT 3884 3888 DAGEY -> RRRRI (IN REF. 1).

FT CONFLICT 3929 3929 A -> V (IN REF. 1).

FT CONFLICT 5134 5134 A -> P (IN REF. 1).

FT CONFLICT 5145 5145 T -> S (IN REF. 1).

FT CONFLICT 5165 5185 G -> A (IN REF. 1).

FT CONFLICT 5199 5199 K -> N (IN REF. 1).

FT CONFLICT 5202 5202 L -> F (IN REF. 1).

FT CONFLICT 5213 5213 F -> L (IN REF. 1).

FT CONFLICT 6178 6178 A -> G (IN REF. 1).

FT CONFLICT 6268 6268 K -> E (IN REF. 1).

SQ SEQUENCE 6632 AA; 731665 MW; 2623EDD62960E89 CRC64;

Query Match 5.1%; Score 174.5; DB 1; Length 6632;  
Best Local Similarity 22.2%; Pred. No. 0.017;  
Matches 127; Conservative 75; Mismatches 200; Indels 171; Gaps 27;

QY 34 GDTVELTCTAQCKSIQFHW-KNSNQIKILNQGSFLTKGSKLNDRADSRSLMDQNF 92  
DB 3934 GETAVLBEKIKSGKPKSPVKKWYKNGEELK-----PS---DRVKIEN--LDDSTQ 3976  
QY 93 PLIINKLIEBDSDFYICGEVDQKEEVQLVFGLTANSDTHLLQGSFLTLIESPPGSSPS 152



```

Db 3977 RLTVNADLDMDVEYRCASNEFGDV-----MSDVTLYKEPQVAFG 4019
Oy 153 VOCRSPRGKNIQGGKTLVSQLELDSDGTCTVLQONOKVEFKIDIVLAFOKASIVY 212
Db 4020 F-----FRELAIQVKEFETAKFECKVSGTSPDKM-----FKQGTPL-- 4057
Oy 213 KKEGQVEFSPFLAFTVEKLTGSGELMWQAERASSKSMITFDLKNKEVSKRYTQDKL 272
Db 4058 -KEDKRVHF-----ESTDDG-----TQRLVEDSKT 4082
Oy 273 QMGKPLPHLTLPLQALPQVASSGNLTALAEKTLGHQVNLVYM--RATLOKRLTCEV 330
Db 4083 D-----DQGNRIEVSNDAGVANSKVPITVVPSETLXIKKGLT-DV 4122
Oy 331 WGPSPKMLSLKLENKEAKV-----SKREKPVWLANPE--AGMMQCLLSDSGQ 377
Db 4123 NVTQGTKLLSVEVGKPKTKYKMYGTETVSSQTKLVQVTESEYKLEISAEKSDIGA 4182
Oy 378 VLESNIKVLPTWSTPVEPKSCDTHTCPPCAPPELLGQSPVFLPPPKDQTLMSRTP 437
Db 4183 YRV-----VLSTDSFVSSESA-----TVTVTKAAEKISLPSF-----KGLADQSVK 4225
Oy 438 VTCVAVDVSHEDPEYKFMVYDGVVHNAAKTRBEOYNSTRVSVLTVLHODWLNKE 497
Db 4226 GTPPLVLEIEIGKPKDVKMYKNGDEIKDGKV---EDLNGKVR---LTIPDFQEKDVG 4278
Oy 498 YKCKSNKALPAPIEKTSKAK-GQRPPOVYT-LPPS---RDELTKNQSVLTCVKGFY 552
Db 4279 YSVTPANBA--GEIF---SKAKNVSAKPEIVSGLVPTVQGEYATNVAVKQPVKG-- 4331
Oy 553 PSDIAVEMESNGQEPNNYKTPPVLDSDGSFFL 585
Db 4332 -----VKWYKNGKEIPDAKTD--NGDGYSYL 4356

RESULT 64
LAC2_RAT STANDARD; PRT; 104 AA.
AC P20767;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 1g lambda-2 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87305594; PubMed=3114047;
RA Steen M.L., Hellman L., Petersen U.;
RT "The immunoglobulin lambda locus in rat consists of two C lambda
RT genes and a single V lambda gene.";
RL Gene 55:75-84(1987).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
DR EMBL; M22521; AAA41420.1; ALT_INT.
DR HSSP; P01842; 2MCG.
DR InterPro; IPR007110; 1g-1like.
DR InterPro; IPR003597; 1g_c1.
DR InterPro; IPR003006; 1g_MHC.
DR Pfam; PF00047; 1g; 1.
DR SMART; SM00407; Tgcl; 1.
DR PROSITE; PS50835; 1g_LIKE; 1.

```

```

DR PROSITE; PS00290; 1g_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1
FT DOMAIN 6 99 IG-LIKE.
FT DISULFID 27 85
FT DISULFID 103 103 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 104 AA; 11318 MW; F087906DE43F7276 CRC64;

Query Match 5.0%; Score 172; DB 1; Length 104;
Best Local Similarity 35.3%; Pred. No. 0.0001;
Matches 36; Conservative 17; Mismatches 45; Indels 4; Gaps 2;

Oy 521 QPRE-POVYTLPPSRDELTKNQSVLTVKGFPSDIAVEMESNGQEPNNYKTPPVLD 579
Db 1 QPKSTPLTLVFPSTTEELQGNKATLVCLISDFPSDVEVAMKANGAPISQGVDTANPTKQ 60
Oy 580 DGSFPLSKLTVDKSRWQGNVFSQSMHEALNHYQKSL 621
Db 61 GNKIYASSFLRLTAQWRSRNSFTCOVTHE--GNVTKSL 99

RESULT 65
PGBM_MOUSE STANDARD; PRT; 3707 AA.
AC Q05793;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=92078153; PubMed=1744087;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogel G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "The complete sequence of perlecan, a basement membrane heparan
RT sulfate proteoglycan, reveals extensive similarity with laminin A
RT chain, low density lipoprotein-receptor, and the neural cell adhesion
RT molecule.";
RL J. Biol. Chem. 266:22939-22947(1991).
RN [2]
RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89034110; PubMed=2972708;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogel G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "Identification of cDNA clones encoding different domains of the
RT basement membrane heparan sulfate proteoglycan.";
RL J. Biol. Chem. 263:16379-16387(1988).
CC -1- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Found in the basement membranes.
CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC -1- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -1- SIMILARITY: Contains 3 laminin IV domains.
CC -1- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 3 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 SEA domain.

```







RT		expressed widely in the adult mouse and during embryogenesis.";
RL	Oncogene.15:691-700(1997).	
CC	-I- FUNCTION:	May be involved as a regulatory protein in the transition of undifferentiated proliferating cells to their differentiated state. May also function as a cell adhesion molecule in a broad spectrum of embryonic and adult tissues.
CC	-I- SUBCELLULAR LOCATION:	Type I membrane protein.
CC	-I- ALTERNATIVE PRODUCTS:	
CC	Event=Alternative splicing; Named isoforms=5;	
CC	Comment=Additional isoforms seem to exist;	
CC	Name=1;	
CC	IsoId=P97798-1; Sequence=Displayed;	
CC	Name=2;	
CC	IsoId=P97798-2; Sequence=VSP_002584;	
CC	Name=3;	
CC	IsoId=P97798-3; Sequence=VSP_002595;	
CC	Note=Expression developmentally regulated;	
CC	Name=4;	
CC	IsoId=P97798-4; Sequence=VSP_002596;	
CC	Note=Expression developmentally regulated;	
CC	Name=5;	
CC	IsoId=P97798-5; Sequence=VSP_002597;	
CC	Note=Expression developmentally regulated;	
CC	-I- TISSUE SPECIFICITY:	Widely expressed.
CC	-I- DEVELOPMENTAL STAGE:	EXPRESSED UBICUOUSLY THROUGHOUT THE MID TO LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5 AND E16.5.
CC	-I- SIMILARITY:	Belongs to the immunoglobulin superfamily. DCC family.
CC	-I- SIMILARITY:	Contains 4 immunoglobulin-like C2-type domains.
CC	-I- SIMILARITY:	Contains 6 fibronectin type III domains.
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; Y09535; CAAT0727.1; -. .	
DR	HSSP; P02751; ITTF.	
DR	MCD; MG1:1097159; Neol.	
DR	InterPro; IPRO08957; FN_III-like.	
DR	InterPro; IPRO03961; FN_III.	
DR	InterPro; IPRO03962; FNIII_subd.	
DR	InterPro; IPRO07110; Ig_Like.	
DR	InterPro; IPRO03598; Ig_c2.	
DR	Pfam; PF00041; fn3; 6.	
DR	Pfam; PF00047; Ig; 4.	
DR	PRINTS; PRN0014; FNTYPEIII.	
DR	SMART; SMO060; FN3; 6.	
DR	SMART; SMO0408; IGc2; 4.	
DR	PROSITE; PS50835; IG_LIKE; 4.	
KW	Cell adhesion; Repeat; Signal; Transmembrane; Immunoglobulin domain;	
KW	Glycoprotein; Alternative splicing.	
FT	SIGNAL	1..36 POTENTIAL.
FT	CHAIN	37..1493 NEOGENIN.
FT	DOMAIN	37..1136 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1137..1157 POTENTIAL.
FT	DOMAIN	1158..1493 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	63..158 IG-LIKE C2-TYPE 1.
FT	DOMAIN	163..249 IG-LIKE C2-TYPE 2.
FT	DOMAIN	254..347 IG-LIKE C2-TYPE 3.
FT	DOMAIN	352..437 IG-LIKE C2-TYPE 4.
FT	DOMAIN	467..564 FIBONECTIN TYPE-III 1.
FT	DOMAIN	567..660 FIBONECTIN TYPE-III 2.
FT	DOMAIN	661..760 FIBONECTIN TYPE-III 3.
FT	DOMAIN	766..860 FIBONECTIN TYPE-III 4.
FT	DOMAIN	861..981 FIBONECTIN TYPE-III 5.



```

FT DOMAIN 982 1083 FIBROBLAST TYPE-III 6.
FT DISULFID 1149 1153 POLY-VAL.
FT DISULFID 85 140 BY SIMILARITY.
FT DISULFID 184 232 BY SIMILARITY.
FT DISULFID 281 331 BY SIMILARITY.
FT DISULFID 373 421 BY SIMILARITY.
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 940 940 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 442 461 Missing (in isoform 2).
FT VARSPLIC 863 878 Missing (in isoform 3).
FT VARSPLIC 1086 1096 Missing (in isoform 4).
FT VARSPLIC 1279 1331 Missing (in isoform 5).
SQ SEQUENCE 1493 AA; 163159 MW; 441DE919D5E17C0E CRC64;

Query Match 4.8%; Score 165.5; DB 1; Length 1493;
Best Local Similarity 21.8%; Pred. No. 0.0088;
Matches 159; Conservative 79; Mismatches 262; Indels 211; Gaps 40;

10 LLLVLTALL--PAA-----TQGNKV-----VLGKKGDTVELT 40
25 LLLLLLTLTLGLRFPASGAATKSGPRROGASVTRTFPEFLVEPDTLSVRSGLVILN 84
41 CRASOKKSIQFMHKNQKNOIKILGNOGSLFKGSKINDADSRSLMDGNFLLIKNL- 99
85 CSAYSESPENIEK-----KQSTP-----LNLESDDRKQLPDGS--LFTSNV 126
100 ---KIEDSDTYIC--EVED---QKEEYQLLVFGL---TANSDFHLQSGSLTLTLES 145
127 HSKMKNPDGEGFYQCVATVNLGTVISRTAKLTVAAGLPRTSQEPSPSVVGSALINCEV 186
146 PEGSSPSV---QCRSP-----RKNKIQQGKTLVSQLELQDSGTWTCTVLQN-----QKK 192
187 NADLVFVFWMEQNRQPLLDLRIVKLPSSG-TLVISNATEGDGLYRCIVESGGPKFSDE 245
193 VEFKI-----DIVVLAFOKASSIVYKKEGEQVEFSPLAFYBEKLTGSGELMWQARAS 247
246 AELKVLQDEBEIVDLVFLMRPSSMKVYTGQ---SAVLPCVVGSLPAPVVRW----- 293
248 SKSWITFDLKNKEVSVKRVTQDPKLOMGKKLPLHLTLPOALPOY---AGSGNLT----- 298
294 -----MKNEEVLDTSESSGRVLVLAGGCLLEISDVTEDDAGTYFCIALNGKNTVEAONE 345
299 LALEAKTGLHQLHOEVLVNRATOLQKNTLCEVWGPSTPKML-----SLKLENKEAKVSR 354
346 LTVQVPPGLKQCPANIVAHESMDIV--FCEVETGKPTFVKKVKNQGDVILPDSNFKIVXE 403
355 E--KPYWVUNPFGKMQCLL-SDSQ-----VLLESNIKVLPTMTSTVEKSCDKTIT 404
404 HNLQVIGLVKSDGEGFYQCIANDVNGAQAQOILIEHNV-AIPT-LPPTSLTSATTDHL 461
405 CPFCPAPELLGGSVFLPPKPKDTLMISRTPEVTCVVVDVSHEDDEVFNMYVQGVENH 464
462 ARAITGP-----LPSAPRD-----VVASLVSTR---IKLTMTPTASDPH 498
465 NAKTKREBOYNSTYRVSVLTVLHODMLNGKYEKKVANKALPABIEKTIIS----- 516
499 GD-----NLTVSVFTKEGVDR-----RVNTSQPGEMQVTIQLMLPATVY 540
517 --KAKGPREPOVYTLPPSRDELTKQVSL-----TCLVKGYRPSPIANWES---NQO 565
541 IFRVMAQNKGSGESSAPLRVE--TOPEVOLPGAPAPNIRAYATSPSITVTWETPLSGNCE 599
566 PENNYKTPTPVLDSDGSFLYSLKLTVDKSRMOGQNVFSCSVNHEALHNHYTKSLSL--- 622

```

```

Db 600 IQ-NRK-----LYWEKGTDE--QDIVSSHSTYINCL-KYTESPRVAV 643
Qy 623 ---SPGLQUDE 630
Db 644 NKRGGVSTOD 654

RESULT 67
LAC3_MOUSE STANDARD; PRT; 104 AA.
ID LAC3_MOUSE
AC P01845;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda-3 chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83014953; PubMed=6812053;
RA Selsing E., Miller J., Wilson R., Storb U.;
RT "Evolution of mouse immunoglobulin lambda genes.";
RN Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).
RN [2]
RP SEQUENCE OF 1-18 (MYELOMA PROTEIN CBPC-49).
RA Breyer R.M., Sauer R.T., Eisen H.N.;
RT "The variable region of mouse lambda-3 chains.";
RN ICN UCLA Symp. Mol. Cell. Biol. 20:105-110(1981).
RN [3]
RP SEQUENCE OF 10-104 (MYELOMA PROTEIN CBPC-49 AND MONOCLONAL ANTIBODY
RX MEDLINE=81223782; PubMed=6165998;
RA Azuma T., Steiner L.A., Eisen H.N.;
RT "Identification of a third type of lambda light chain in mouse
RN Proc. Natl. Acad. Sci. U.S.A. 78:569-573(1981).
RN [4]
RP -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC
DR EMBL; J00585; AAB59670.1; -
DR PIR; B93922; L3MS.
DR HSP; P01842; 2MCG.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IGC1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 1 99
FT DISULFID 27 85 IG-LIKE.
FT DISULFID 103 103 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 104 AA; 11371 MW; 83CEBCDD4A348EF1 CRC64;

Query Match 4.8%; Score 163; DB 1; Length 104;
Best Local Similarity 34.3%; Pred. No. 0.0038;
Matches 35; Conservative 16; Mismatches 47; Indels 4; Gaps 2;

521 QPRE-POVYTLPPSRDELTKQVSLTCLVKGYRPSPIANWESNQGPENNYKTPTPVLD 579
1 QPKSPFTLTMPSPSEELQENKATLVCLISNFPSPGCVVANKANGTPTTQGVDTSPNPK 60

```



```

Oy 580 DGSFPLYSKLTVDKSRMOQGNVFCSCVMHEALHNHYTOKSL 621
Db 61 DNKYMASFHLHTSDQWHSNHSFTCOVH---GDTYKSLIS 99

RESULT 68
ID ILL1_HUMAN STANDARD; PRT; 213 AA.
AC P15814.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Immunoglobulin lambda-like polypeptide 1 precursor (Immunoglobulin-
DE related 14.1 protein) (Immunoglobulin omega polypeptide) (Lambda 5)
DE (CD179b antigen).
GN IGLL1 OR IGL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89315835; PubMed=2501791;
RA Hollis G.F., Evans R.J., Stafford-Hollis J.M., Kozsmeier S.J.,
RA McKearn J.P.;
RT "Immunoglobulin lambda light-chain-related genes 14.1 and 16.1 are
RT expressed in pre-B cells and may encode the human immunoglobulin
RT omega light-chain protein."
RL Proc. Natl. Acad. Sci. U.S.A. 86:5552-5556(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Lymphoid;
RA MEDLINE=91108327; PubMed=1703205;
RA Evans R.J., Hollis G.F.;
RT "Genomic structure of the human Ig lambda 1 gene suggests that it may
RT be expressed as an Ig lambda 14.1-like protein or as a canonical B
RT cell Ig lambda light chain: implications for Ig lambda gene
RT evolution."
RL J. Exp. Med. 173:305-311(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Testis;
RA MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Schein C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marulanda K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M., Soares M.B., Bonaldi U.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wozney K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
RA Butlerfield A.S., Schin J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SUBMIT: Associates non-covalently with VPREB1.
CC -1- TISSUE SPECIFICITY: EXPRESSED ONLY IN PRE-B-CELLS AND A SPECIAL
CC B-CELL LINE (WHICH IS SURFACE IG NEGATIVE).
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC -1- DATABASE: NAME=PROW; NOTE=PROW 1:64-67(2000);
CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/696419174_g.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M27749; AAA36100.1; -.
DR EMBL; M34513; AAA36096.1; -.
DR EMBL; M34511; AAA36096.1; JOINED.
DR EMBL; M34512; AAA36096.1; JOINED.
DR EMBL; BC012293; AAH12293.1; -.
DR PIR; A33911; A33911.
DR HSSP; P01842; 7FAB.
DR Genew; HGNC:5870; IGLL1.
DR MIM; 146770; -.
DR GO; GO:0016020; C:membrane; NAS.
DR GO; GO:0003793; F:defense/immunity protein activity; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IGL1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Antigen; Signal; Immunoglobulin domain.
FT SIGNAL 1 37
FT CHAIN 38 213
FT DOMAIN 97 108
FT DOMAIN 109 213
FT DOMAIN 109 213
SQ SEQUENCE 213 AA; 22963 MW; 9133A742B943C79 CRC64;
Query Match 4.8%; Score 162.5; DB 1; Length 213;
Best Local Similarity 30.5%; Pred. No. 0.0011;
Matches 36; Conservative 21; Mismatches 44; Indels 17; Gaps 3;
Oy 521 QPR-EPQVYTLPPSRDELTKNOVSLTCTVKGFPYSDIAYMESNGOP-ENNYKTPPLVD 578
Db 109 QPKATPSVTLPPSSSELTQNKATLVCMDFPGILITVWKADGITITGCVETTTTSKO 168
Oy 579 SDGSFPLYSKLTVDKSRMOQGNVFCSCVMHEALHNHYTOKSLSPGLQDETCAEQ 636
Db 169 SNKYMASFHLHTSDQWHSNHSFTCOVH---GDTYKSLIS 211

RESULT 69
ID CD22_HUMAN STANDARD; PRT; 847 AA.
AC P20273; Q95699; Q95701; Q95702; Q95703; Q01665; Q92872; Q92873;
AC Q9UQA7; Q9UQA8; Q9UQA9; Q9UQA0; Q9Y2A6;
DT 01-FEB-1991 (Rel. 17, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE B-cell receptor CD22 precursor (Lew-14) (B-lymphocyte cell adhesion
DE molecule) (BL-CAM) (Siglec-2).
GN CD22.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM CD22-BETA), AND VARIANT HIS-639.
RX TISSUE=Tonsil;
RA MEDLINE=91086838; PubMed=1985119;
RA Wilson G.L., Fox C.H., Fauci A.S., Kehr J.H.;
RT "cDNA cloning of the B cell membrane protein CD22: a mediator of B-B
RT cell interactions."
RL J. Exp. Med. 173:137-146(1991).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS CD22-ALPHA AND CD22-BETA).
RX MEDLINE=93267103; PubMed=8496602;

```







FT	DOMAIN	242	326		IG-LIKE C2-TYPE 2.
FT	DOMAIN	331	416		IG-LIKE C2-TYPE 3.
FT	DOMAIN	419	500		IG-LIKE C2-TYPE 4.
FT	DOMAIN	505	582		IG-LIKE C2-TYPE 5.
FT	DOMAIN	593	676		IG-LIKE C2-TYPE 6.
FT	SITE	760	765		ITIM MOTIF 1.
FT	SITE	794	799		ITIM MOTIF 2.
FT	SITE	820	825		ITIM MOTIF 3.
FT	SITE	840	845		ITIM MOTIF 4.
FT	DISULFID	39	167		BY SIMILARITY.
FT	DISULFID	44	102		BY SIMILARITY.
FT	DISULFID	161	219		BY SIMILARITY.
FT	DISULFID	265	309		BY SIMILARITY.
FT	DISULFID	353	396		BY SIMILARITY.
FT	DISULFID	442	484		BY SIMILARITY.
FT	DISULFID	529	571		BY SIMILARITY.
FT	DISULFID	616	659		BY SIMILARITY.
FT	MOD_RES	762	762		PHOSPHORYLATION (INVOLVED IN BINDING TO SYK)
FT	MOD_RES	807	807		(BY SIMILARITY).
FT	MOD_RES	822	822		PHOSPHORYLATION (INVOLVED IN BINDING TO GRB2) (BY SIMILARITY).
FT	MOD_RES	842	842		PHOSPHORYLATION (INVOLVED IN BINDING TO SYK) (BY SIMILARITY).
FT	CARBOHYD	67	67		PHOSPHORYLATION (INVOLVED IN BINDING TO SYK, PLCG2 AND PI3K1/PI3K2) (BY SIMILARITY). N-LINKED (GLCNAC. . .) (POTENTIAL).
Query Match Similarity 4.8%; Score 162.5; DB 1; Length 847;					
Beet Local Similarity 19.4%; Pred. No. 0.0065;					
Matches 132; Conservative 103; Mismatches 229; Indels 217; Gaps 36					
OY	LIINKLKIEDSDTYICE-----VEDQKEEVQLVFGLTANSDPTL	133			
Dd	11 LVLETLASDSBSKWVEHPETLYAMEGACWLPCTRYALDGBESFILFHNEPYKNNTSK	70			
OY	134 LOGSLSLTLE--SPGGSSPVQCRPPGKNIOGGKTLSVSQELDODSGTWTCTVLONOK	191			
Dd	71 FDGTRLYESTKDGVPSQKRQVQFGDKNKNC---TLSHVHLNDSSQ---LGIRMES	123			
OY	192 KVERKIDIVL-----AFQKASSIYKKEGEQQE-----FSFLAFTVEKLT	233			
Dd	124 KTEKMMEIRHNLNVSERPPEPHIQLPPEI---QESEVTLTCLLNPSCYGPV-----	172			
OY	234 GSGETLMQAERASSSKSWTFDLRKKEYSVKRVTDPKLOMGKPLRHLLTPQALPOYAG	293			
Dd	173 ---QLQHLEGVPRQAAVT-----STSLIKSVFRSELKRS-----POMSH	212			
OY	294 SGNL-TLALAKTKCL--HOENVLVMMRATOLQ-----KNLTCEVMGPTS	335			
Dd	213 HGKITVTQLOADGKFLENDFYQLNVHKTPKLEIKVTPSDAIRBGDSVTWMCEV-SSSN	271			
OY	336 PKM-----LSIKLENKAUKSKREKVPWVNLPERGAMOCCLSD-----SCQVILE	381			
Dd	272 PEYTTWSMLKDGTSILKQN-TFTLLNR-----VTXDOGAKKCVCQVNSDVGRSESEVFLO	326			
OY	382 -----SNIKVLP-----TW-----SRPVEKSGCDKHHT-----	404			
Dd	327 VOYAPEBSTVOILHSFAVGSOVELCMSLANPLFTNYTWNHGEMOQRTEKVHIPIKI	386			
OY	405 -----CPPCAPBELLG-----GPSVFL----FPFKEDTLMTSRTP-----EYTCVVVDVSH	447			
Dd	387 LPMHAGTVSCVAENITGTGQRGPAELDVOYEPKKTVTYIQNPMPHIRBEDPTTSLCANNYS	446			
OY	448 EDPEY-KNNAYVDGYEVFNAAKTPREEQNSTYRVVSVLTJLHQDLNGEKYEYCKCVSNK-	505			
Dd	447 SNPSTRTRYEW-----KPGHAMEEPS-----LGVLIKIONVGMDN-TTIACAACSMS	490			
OY	506 -ALPAPIEKITSKAGQPREQPVYLLPSPRDELTKNOVSTLCUVAIGFYPSDLAVMEENG	564			

```

Db      491 CSMA$PVALN0YA---PRDVRVKRIKPLSEIHSGNSVSLQCFSSHHKXVQFFMEKNG 547
Oy      565 -----QPENNYKTTTPVLVSDSGSFPLYSKLTVDKSRWQGNVFGSCVMEALHNHTYQS 619
Db      548 RLKGESQTLNPDSPISF--EDAGSY-----SCWVNSISGQTFASKAMTLEVLVAPRR 595
Oy      620 L--SLSPGLQDDE-----TC 632
Db      596 LRVSMSPGDQVMEGK$ATLTC 616

RESULT 70
VCAM1 HUMAN
ID _VCAM1 HUMAN STANDARD; PRT; 739 AA.
AC P19320;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vascular cell adhesion protein 1 precursor (V-CAM 1) (CD106 antigen)
DE (INCAM-100).
GN VCAM1 OR LICAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBL_TaxID=9606;
RX [1]
RN RP SEQUENCE FROM N.A.
RC TISSUE=Umbilical vein;
RX MEDLINE=91016951; PubMed=169207;
RA Polte T., Newman W., Gopal T.V.;
RL "Full length vascular cell adhesion molecule 1 (VCAM-1).";
RN Nucleic Acids Res. 18:5901-5901 (1990).
RN [2]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=90090619; PubMed=268898;
RA Osborn L., Hession C., Tizard R., Vassallo C., Luhowskyj S.,
RA Chi-Rosso G., Lobb R.;
RT "Direct expression cloning of vascular cell adhesion molecule 1, a
RT cytokine-induced endothelial protein that binds to lymphocytes.";
RL Cell 59:1203-1211 (1989).
RN [3]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=91352090; PubMed=1715583;
RA Cybulsky M.I., Fries J.W.U., Williams A.J., Sultan P., Eddy R.,
RA Byers M., Shows T., Gimbrone M.A. Jr., Collins T.;
RT "Gene structure, chromosomal location, and basis for alternative mRNA
RT splicing of the human VCAM1 gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:7859-7863 (1991).
RN [4]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=91201302; PubMed=1707873;
RA Hession C., Tizard R., Vassallo C., Schiffer S.B., Goff D., Moy P.,
RA Chi-Rosso G., Luhowskyj S., Lobb R., Osborn L.;
RT "Cloning of an alternate form of vascular cell adhesion molecule-1
RT (VCAM1).";
RL J. Biol. Chem. 266:6682-6685 (1991).
RN [5]
RN RP SEQUENCE FROM N.A., AND VARIANTS PHE-318; ALA-384; ALA-413 AND
RP LEU-716.
RA Rieder M.J., Armel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,
RA Rajkumar N.R., Toch E.U., Yi Q., Nickerson D.A.;
RL Submitted (Aug-2002) to the EMBL/Genbank/DBJ databases.
RN [6]
RN RP SEQUENCE FROM N.A. (ISOFORM LONG).
RC TISSUE=Retinal pigment epithelium;
RX MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Aleschul S.F., Zeeberg B., Butow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.B., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

```



RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Raza S.A., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
RA Villalón D.K., Muzny D.M., Sodegren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.M., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smalhus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Maira M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[7]  
RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 25-226.  
RP MEDLINE:95147978; PubMed:7531291;  
RX Jones E.Y., Harlow K., Bottomley M.J., Robinson R.C., Driscoll P.C.,  
RA Edwards R.M., Clements J.M., Dudgeon T.J., Stuart D.I.,  
RT "Crystal structure of an integrin-binding fragment of vascular cell  
RT adhesion molecule-1 at 1.8-A resolution.";  
RL Nature 373:539-544(1995).  
[8]  
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.  
RP MEDLINE:95296382; PubMed:753925;  
RX Wang J.-H., Pepinsky R.B., Stehle T., Liu J.-H., Karpusas M.,  
RA Browning B., Osborn L.,  
RT "The crystal structure of an N-terminal two-domain fragment of  
RT vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on  
RT the domain 1 C-D loop can inhibit VCAM-1-alpha 4 integrin  
RT interaction.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:5714-5718(1995).  
[9]  
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.  
RP Wang J.-H., Stehle T., Pepinsky R.B., Liu J.-H., Karpusas M.,  
RA Osborn L.,  
RT "Structure of a functional fragment of VCAM-1 refined at 1.9-A  
RT resolution.";  
RL Acta Crystallogr. D 52:369-379(1996).  
CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION  
CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1  
CC INTEGRIN VIA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL  
CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A  
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE  
CC EMIGRATION TO SITES OF INFLAMMATION.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Comment=Additional isoforms seem to exist;  
CC Name=Long;  
CC IsoId=PI9320-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoId=PI9320-2; Sequence=VSP\_002580.  
CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as  
CC well as on macrophage-like and dendritic cell types in both normal  
CC and inflamed tissue.  
CC -1- INDUCTION: By cytokines (e.g. IL-1, TNF-alpha).  
CC -1- PTM: Sialoglycoprotein.  
CC -1- DISEASE: May play an important role in the genesis of  
CC atherosclerosis and rheumatoid arthritis.  
CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.  
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD106 entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd106.htm".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; X53051; CAA57218.1; -  
CC EMBL; M30257; AAA51917.1; ALT\_TERM.  
CC DR

DR EMBL; M73255; AAA61270.1; -  
DR EMBL; M60335; AAA61269.1; -  
DR EMBL; AF536818; AAA963190.1; -  
DR EMBL; BC01276; AAA1276.1; -  
DR PIR; A41288; A41288.  
DR PIR; B41288; B41288.  
DR PDB; 1VCA; 15-SEP-95.  
DR PDB; 1VSC; 20-JUN-96.  
DR PDB; 11I9; 07-NOV-01.  
DR Genew; HGNC:12663; VCAM1.  
DR MIM; 192225; -  
DR InterPro; IPR003987; ICAM VCAM-1.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG C2.  
DR InterPro; IPR003989; VCAM-1.  
DR Pfam; PF00047; Ig; 6.  
DR PRINTS; PR01472; ICAMVCAM1.  
DR PRINTS; PR01474; VCAM1.  
DR SMART; SM00408; IGC2; 3.  
DR PROSITE; PS50835; IG-LIKE; 5.  
KW Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;  
KW Repeat; Signal; Alternative splicing; Polymorphism; 3D-structure.  
FT SIGNAL 1 24  
FT CHAIN 25 739  
FT DOMAIN 25 698  
FT TRANSMEM 699 720  
FT DOMAIN 721 739  
FT DOMAIN 25 105  
FT DOMAIN 109 212  
FT DOMAIN 223 309  
FT DOMAIN 312 399  
FT DOMAIN 408 506  
FT DOMAIN 511 595  
FT DOMAIN 600 684  
FT DISULFID 47 95  
FT DISULFID 52 99  
FT DISULFID 137 195  
FT CARBOHYD 273 273  
FT CARBOHYD 365 365  
FT CARBOHYD 417 417  
FT CARBOHYD 463 463  
FT CARBOHYD 531 531  
FT CARBOHYD 561 561  
FT VARPELIC 310 402  
FT VARIANT 318 318  
FT VARIANT 384 384  
FT VARIANT 413 413  
FT VARIANT 716 716  
FT STRAND 26 30  
FT STRAND 34 38  
FT TURN 39 40  
FT STRAND 43 50  
FT STRAND 56 61  
FT TURN 62 63  
FT STRAND 70 74  
FT TURN 75 76  
FT STRAND 77 82  
FT HELIX 87 89  
FT STRAND 92 99  
FT TURN 100 101  
FT STRAND 102 114  
FT STRAND 120 123  
Query Match 4.7%; Score 161.5; DB 1; Length 739;  
Best Local Similarity 21.5%; Pred. No. 0.0063;  
Matches 118; Conservative 91; Mismatches 217; Indels 123; Gaps 27;  
QY 8 RHLLVLQALLPAT--QNKVVLGKGGDVELTCTASQKSKIQFHKNKSNQIKILNQ 65



Ds	211	ROAVKELQYVISPKNTVIVISNPGTKIQBGSVYTMCSSEGLPAPITFW-----	258
Qy	66	GSEFLTKGPTSLANDRADSRSLMDOGNFPILIKNLTIESDPTICE---VEDQKEEYQL	121
Ds	259	-----SKKLDNGNLQHLL---SGNATLTTLAMEMESGIVYCEGVNLIGKRKEVELI	307
Qy	122	V----FGLTNSDTHLLQ---GQSLTTLTLESPPGSSPSVQCSPPKAKIQG-----GKT	168
Ds	308	VOEKPFVVEISPEPRITAAQIGDSVNLTCSSVMGCEBSPFSMRTQIDSPLSGKVRSEGTNST	367
Qy	169	LSVSOLELDOSSGWTCTVQLONQKVEPFKIDIVLAFQKASSIVK---KEGEYEFs---	222
Ds	368	LTLSPVSEFENHSYLTCTVCGHKLEKGIQVLYISFPDPELEMSGGLVNSGSVYSCV	427
Qy	223	---FPLAFVTEKLTSGELIMWOAERASSSKSWITF---DLKKEVSVKRVTDOPKIQ-	273
Ds	428	PSVYPPDLRLEIELLKKEITL-----ENIEFLDITMKSILENKSLEMTITPIED	476
Qy	274	MKKKL---PLHLTFPOLPYQVSGNLTALAEATGKLHQLHENVLVNARATOLQK---N	335
Ds	477	TGKALVQQAALHIDHMEFEPKORQs---TQTLVYNVAP--RDTVLVSPSSIIIEGSSVN	531
Qy	326	LTECWGPTSPKMLSLKLENKAYSKREKRWVLN---PEAGMQCL-LSBGGQVLE	381
Ds	532	MTLGSGFPAPKTLWRSQLPNSBELQPLSENATLTLLSTYRMBDSGVYLCGGINQGRSRKE	591
Qy	382	SNIKVLPTWSTPVEPSPCKTHTCPCPAPBELLGSPVFLPPP---KPOTLMISRT---	435
Ds	592	VELLI-----QYTPDKLT-----APFSSEVGEQTVIISCTCGN	627
Qy	436	-PEVTCVVVDVSHEDVEKFNMYVDGV-EVNAKTKPRE---EQYNSTYRVVSVLTVLH	489
Ds	628	VPE-TWIIKKKAETGDTVLK-SIDGAVYIRKAQKDGAYECESKVKVQSGLRSLTDV	685
Qy	490	QDWLNGKEY	498
Ds	686	QGRENNKDY	694

RESULT 71

LAC2\_MOUSE

ID LAC2\_MOUSE STANDARD; PRT; 104 AA.

AC P01844;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE ig lambda-2 chain C region.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxId=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=83014953; PubMed=6812053;

RA Seising E., Miller J., Wilson R., Storb U.;

RT "Evolution of mouse immunoglobulin lambda genes.";

RL Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).

RN [2]

RP SEQUENCE FROM N.A.

RA MEDLINE=82274221; PubMed=6287422;

RA Wu G.E., Govindji N., Hozumi N., Murialdo H.;

RT "Nucleotide sequence of a chromosomal rearranged lambda 2

RL immunoglobulin gene of mouse.";

RN [3]

RP SEQUENCE FROM N.A. (MOPC 315).

RA MEDLINE=82220143; PubMed=6283385;

RA Borchwell A.L.M., Paskind M., Reith M., Imanishi-Kari T., Rajewsky K.,

RT "Somatic variants of murine immunoglobulin lambda light chains.";

RL Nature 298:380-382(1982).

RN [4]

```

RP SEQUENCE (MOPC 315).
RX MEDLINE=74048693; PubMed=4760498;
RX Dusan E.S., Bradshaw R.A., Simme E.S., Eisen H.N.;
RT "Amino acid sequence of the light chain of a mouse myeloma protein
RT (MOPC-315).";
RL Biochemistry 12:5400-5416(1973).
RN [5]
RP SEQUENCE OF 66-104 (MOPC 315), AND REVISIONS.
RX MEDLINE=81223782; PubMed=6165998;
RX Azuma T., Steiner L.A., Eisen H.N.;
RT "Identification of a third type of lambda light chain in mouse
RT immunoglobulins.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:569-573(1981).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, J00595; AAA39151.1; -.
DR PIR, C93922; LZMS.
DR HSSP, P01842; 2MCG.
DR InterPro: IPR007110, IG-1ike.
DR InterPro: IPR003597, IG_C1.
DR InterPro: IPR003006, IG_MHC.
DR Pfam, PF00047, Ig_1.
DR SMART, SM00407, IgC1, 1.
DR PROSITE, PS50835, IG_LIKE, 1.
DR PROSITE, PS00290, IG_MHC, 1.
DR Immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1 1
FT DOMAIN 6 99 IG-LIKE.
FT DISULFD 27 85
FT DISULFD 103 103 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 104 AA; 11254 MW; CE4B67B868862D3 CRC64;

Query Match 4.7%; Score 161; DB 1; Length 104;
Best Local Similarity 34.3%; Pred. No. 0.00051;
Matches 35; Conservative 15; Mismatches 48; Indels 4; Gaps 2;

Oy 521 QPRE-PCVYVTTTPPSPDELTQKQVSLTCLVKKFYSDSLAVENSGOPENNYKTTTPVLDS 579
Db 1 QPKSTPTITVPPSESELKEMKATLVCLISFSGVYVANKAKGPTITQGVDSNPRIKE 60

Oy 580 DGSFFLYSKLTVDKSRMOQGQNVFSCSVMEALHNHYTKSKIS 621
Db 61 GNKEWASSFLHILTSQDWRSHNSFTCQVTHE---GDVTEKSKIS 99

RESULT 72
HB2D_PIG STANDARD; PRT; 258 AA.
AC P15983;
ID HB2D_PIG
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SLA class II histocompatibility antigen, DQ haplotype D beta chain
DE precursor.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90361905; PubMed=2391424;
RX Guettersson K., Leguenn C., Hirsch F., Germana S., Pratt K.,
RA Sachs D.H.;
RT "Class II genes of miniature swine. IV. Characterization and
RT expression of two allelic class II DQB CDNA clones.";
```



```

RL J. Immunol. 145:1946-1951(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M1498; AAA31085.1; .
DR HSSP, P13760; 2SEB.
DR InterPro: IPR007110; IG_1like.
DR InterPro: IPR003597; IG_c1.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR000353; MHC_II_beta.
DR Pfam: PFO0047; IG_1.
DR Pfam: PFO0969; MHC_II_beta; 1.
DR ProDom: PDO00328; MHC_II_beta; 1.
DR SMART: SM00407; IGc1_1.
DR PROSITE: PSS0835; IG_LIKE; 1.
DR PROSITE: PS00290; IG_MHC; 1.
DR MHC II; Transmembrane; Glycoprotein; Signal.
KW MHC II;
FT CHAIN 1 31
FT SIGNAL 1 258
FT DOMAIN 32 123 SLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 124 217 DO HAPLOTYPE D BETA CHAIN.
FT DOMAIN 218 227 EXTRACELLULAR BETA-1.
FT TRANSMEM 228 248 EXTRACELLULAR BETA-2.
FT DOMAIN 249 258 CONNECTING PEPTIDE.
FT DISULFID 44 108 CYTOPLASMIC TAIL.
FT DISULFID 146 202 BY SIMILARITY.
FT CARBOHYD 48 48 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 258 AA; 29262 MW; E3AC75110AED47C3 CRC64;
Query Match 4.7%; Score 161; DB 1; Length 258;
Best Local Similarity 28.4%; Pred. No. 0.0017;
Matches 46; Conservative 28; Mismatches 80; Indels 8; Gaps 4;
QY 456 WYVGVENNAKTPREQYNSTYRVSVLYLVHQMILNGKEYKCKVSNKALPAPIEKTI 515
DB 54 WSDVDR-INYQSEFLAFDSMDGCVRAVTPVLRPDADNYLNGQKAELEDTYCKINY 112
QY 516 SKAKG---QPREQVYTLPPSRDELTKNOVSLTCLVKGYYPSDIAVWESNGQPER-NYK 571
DB 113 QIEGGTLORVQPTVTISPSKAEALHNHNLVCAVTDYPSQVQKQPMRNGQETAGVV 172
QY 572 TTPPVLDSDGSPFLYSKLTVDKSRMQGQNVFSGSVHGAALHN 613
DB 173 STPLRNGD--WTYQVLMLENNLQRGDVGVTCTCRVHSSLSQS 211

```

```

RT RT Cloning of the complete gene for carcinoembryonic antigen: analysis
RT of its promoter indicates a region conveying cell type-specific
RT expression.";
RT Mol. Cell. Biol. 10:2738-2748(1990).
RN (2)
RN SEQUENCE FROM N.A.
RX MEDLINE=89122014; PubMed=3670312;
RA Beauchemin N., Benchimol S., Cournoyer D., Fuks A., Stramers C.P.;
RT "Isolation and characterization of full-length functional cDNA clones
RT for human carcinoembryonic antigen.";
RT Mol. Cell. Biol. 7:3221-3230(1987).
RN (3)
RN SEQUENCE FROM N.A.
RX MEDLINE=89122014; PubMed=3320478;
RA Barnett T., Goebel S.J., Nothdurft M.A., Elting J.J.;
RT "Carcinoembryonic antigen family: characterization of cDNAs coding
RT for NCA and CEA and suggestion of nonrandom sequence variation in
RT their conserved loop-domains.";
RT Genomics 3:59-66(1988).
RN (4)
RN SEQUENCE OF 5-702 FROM N.A.
RX MEDLINE=87128144; PubMed=3814146;
RA Oikawa S., Nakazato H., Kosaki G.;
RT "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT from cDNA sequence.";
RT Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN (5)
RN SEQUENCE OF 331-702 FROM N.A.
RX MEDLINE=87204247; PubMed=3033671;
RA Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT "Isolation and characterization of cDNA clones encoding the human
RT carcinoembryonic antigen reveal a highly conserved repeating
RT structure.";
RT Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
RN (6)
RN SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC - TISSUE SPECIFICITY: Found in adenocarcinomas of endodermally
CC derived digestive system epithelium and fetal colon.
CC - PTM: COMPLEX IMMUNOREACTIVE GLYCOPROTEIN WITH A MW OF 180 kDa
CC COMPRISING 60% CARBOHYDRATE.
CC - SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC - SIMILARITY: Contains 7 immunoglobulin-like domains.
CC - DATABASE: NAME=PROW; NOTE=CD guide CD66 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd66c.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M17303; AAB59513.1; .
DR EMBL, M59262; AAA62835.1; ALT SEQ.
DR EMBL, M59265; AAA62835.1; JOINED.
DR EMBL, M59256; AAA62835.1; JOINED.
DR EMBL, M59257; AAA62835.1; JOINED.
DR EMBL, M59258; AAA62835.1; JOINED.
DR EMBL, M59259; AAA62835.1; JOINED.
DR EMBL, M59260; AAA62835.1; JOINED.
DR EMBL, M59261; AAA62835.1; JOINED.
DR EMBL, M59709; -; NOT_ANNOTATED_CDS.
DR EMBL, M59710; -; NOT_ANNOTATED_CDS.
DR EMBL, M29540; AAA51967.1; .
DR EMBL, X16455; CAA34474.1; .
DR EMBL, M15042; AAA51963.1; .
DR EMBL, M16234; AAA51972.1; .
DR PIR, A36319; A36319.
DR PDB, 1E07; 04-JUL-00.
DR GeneW; HGNC:1817; CEACAM5.
DR MIM; 114890; .
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR InterPro; IPR007110; IG_1like.

```



```

DR Pfam; PF00047; Ig; 6.
DR PROSITE; PS50835; IG_LIKE; 6.
KW Immunoglobulin domain; Lipoprotein; GPI-anchor;
KW Membrane; Signal; Repeat; 3d-structure.
FT SIGNAL 1 34
FT CHAIN 35 685
FT PROPEP 686 702
FT DOMAIN 35 144
FT DOMAIN 146 237
FT DOMAIN 238 322
FT DOMAIN 324 415
FT DOMAIN 416 498
FT DOMAIN 502 593
FT DOMAIN 594 677
FT LIPID 685 685
FT CARBOHYD 104 104
FT CARBOHYD 115 115
FT CARBOHYD 152 152
FT CARBOHYD 182 182
FT CARBOHYD 197 197
FT CARBOHYD 204 204
FT CARBOHYD 208 208
FT CARBOHYD 246 246
FT CARBOHYD 256 256
FT CARBOHYD 274 274
FT CARBOHYD 288 288
FT CARBOHYD 292 292
FT CARBOHYD 309 309
FT CARBOHYD 330 330
FT CARBOHYD 351 351
FT CARBOHYD 360 360
FT CARBOHYD 375 375
FT CARBOHYD 432 432
FT CARBOHYD 466 466
FT CARBOHYD 480 480
FT CARBOHYD 508 508
FT CARBOHYD 529 529
FT CARBOHYD 553 553
FT CARBOHYD 560 560
FT CARBOHYD 580 580
FT CARBOHYD 612 612
FT CARBOHYD 650 650
FT CARBOHYD 665 665
FT CONFLICT 320 320
SQ SEQUENCE 702 AA; 76795 MW; 6299AE26CDDDB5C CRC64;

Query Match 4.7%; Score 161; DB 1; Length 702;
Best Local Similarity 17.7%; Pred. No. 0.0064;
Matches 119; Conservative 108; Mismatches 255; Indels 190; Gaps 32;

OY 20 PATQGNKVVVGGKGVDELCTASQKSIQFHMKNISQIKILNQSGFLTKGSKLNDK 79
DB 147 PSISNSKPEVDK-DAVAFTCEPETOATYLMVNNOSLPV----- 187
OY 80 ADSRRSLWDOGNFPLIKLKIEDSDYICEVED-----QKEEYQL-LVFGLTANS--- 129
DB 188 --SPRLQISNGNRITLTFNVRNDPTASYKCEQNPVARSRDSYILNVLPDAPTISPL 245
OY 130 DTHLQGSGLTTLTSPGSSPSVQCR-SPRGKNIQSGKTLVSQLEQDSGTTCTVQL 188
DB 246 NTSYRSNGNLSLSCA--ASNPPQVYSFVNGTQOSTQELFIRITVNNNGSYTCQAHN 303
OY 189 NQKIV-EFKIDIVVLAFOKASSIYKKEGQEVESFPLAFVE-KLTGSGELMGAERAS 246
DB 304 SDTGLNRTTITVTVYAEPPFITSNNSNVEDEDAVALTCEPIQNTYLLM----- 357
OY 247 SSKSMITFDLKNKEVSVRVTQDPQLQWKKLPLHLPLQALPOYAGSGNLTLLAEKATG 306
DB 358 -----VNNQSLPV-----SPRLQL-----SNDNRTLLTLLSVTR 385
OY 307 KLRHENVLVKRAQLQKNTLCE-----VNGPSP-----KMLSLKL 344

```

```

DB 386 -----NDVGPYECGIQNELSDHSDPVLNLYGPDPTISBYTYRPGVNLISLCHA 439
OY 345 ENK-----BAKYSKREKPPVVLN---PEAGMQCLSDSGQVLLESNIKVLPTMST 392
DB 440 ASNPAPQSWLIDGNITQHQDELFIISNTEKNSGLYTCQANNSAGSRITVKTI-TVSA 498
OY 393 PVEBKSCDKTHCTCPAPPELLGPPSVLFPPPKKDTLMTISRPEVTCVVVDVSHDEPEV 452
DB 499 EL-----PKPISNSNSK---PVEDKDAVAFTCEPEA-----QNT 530
OY 453 KFMVYDGVENHAKTPREEOYNSTRVYVSLTVLHQDLNKGVEYCKVSNKALPAPIE 512
DB 531 TYLMVWVG---QSLPVSRLQLSNGN-RTLTLFVTRND---AAAYVCGIQL----- 575
OY 513 KTSKAGQPREPOV-----YTLPPSRDELTFKQVSLTCLVKGFPSPDIAMWESNQ 565
DB 576 -SVANSRSDVTLIDVLYGPPTPIISPDSSYLSGANLNLSC-HSASNPSP-QYSWRINGI 632
OY 566 PENNYKTPPVLDSGFFLYSKLTYDKSRWQGNVFSGVMEAL-HNHYTKSLSL-- 622
DB 633 PQGHTQV-----LFLAKITPNN-----GTVACFVSNLATGRNNSIVKSIITVSA 676
OY 623 ---SPGLQDET 631
DB 677 SGTPGSLAGAT 688

RESULT 74
LAC_HUMAN STANDARD; PRT; 105 AA.
ID_P01842; P80423;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig lambda chain C regions.
GN IGLC1 AND IGLC2 AND IGLC3.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE (BENCE-JONES PROTEIN SH).
RX MEDLINE=70166723; PubMed=4909564;
RA Tilani K., Wikler M., Shinoda T., Putnam F.W.;
RT "The amino acid sequence of a lambda type Bence-Jones protein. 3. The
RT complete amino acid sequence and the location of the disulfide
RT bridges.";
RL J. Biol. Chem. 245:2171-2176(1970).
[2]
RP SEQUENCE (BENCE-JONES PROTEIN X), AND DISULFIDE BONDS.
RX MEDLINE=69088380; PubMed=4883841;
RA Milstein C., Clegg J.B., Jarvis J.M.;
RT "Immunoglobulin lambda-chains. The complete amino acid sequence of a
RT Bence-Jones protein.";
RL Biochem. J. 110:631-652(1968).
[3]
RP SEQUENCE (BENCE-JONES PROTEIN NIG-64).
RX MEDLINE=63186114; PubMed=6404900;
RA Kametani F., Takayasu T., Suzuki S., Shinoda T., Okuyama T.,
RA Shimizu A.;
RT "Comparative studies on the structure of the light chains of human
RT immunoglobulins. IV. Assignment of a subgroup.";
RL J. Biochem. 93:421-429(1983).
[4]
RP SEQUENCE (BENCE-JONES PROTEIN KERN).
RX MEDLINE=71150336; PubMed=5549568;
RA Pomeroy H., Hess M., Hilschmann N.;
RT "Structural rule of antibodies. Primary structure of a monoclonal
RT immunoglobulin-L-chain of the lambda type, subgroup IV (Bence-Jones-
RT protein Kern). V. The complete amino acid sequence and its genetic
RT interpretation.";
RL Hoppe-Seyler's Z. Physiol. Chem. 352:247-266(1971).
[5]

```



RP PARTIAL SEQUENCE (MYELOMA PROTEIN NEMM).  
RX MEDLINE=74109253; PubMed=4814727;  
RA Chen B.L., Poljak R.J.;  
RT "Amino acid sequence of the (lambda) light chain of a human myeloma  
RL immunoglobulin (IgG New).";  
RT Biochemistry 13:1295-1302(1974).  
RN [6]  
RP SEQUENCE (DOT).  
RX MEDLINE=95255298; PubMed=7737190;  
RA Scoppini M., Bellotti V., Negri A., Merlini G., Garver F., Ferri G.;  
RT "Characterization of the two unique human anti-Flavlin monoclonal  
RL immunoglobulins.";  
RL Eur. J. Biochem. 228:886-893(1995).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF NEMM.  
RX MEDLINE=75046825; PubMed=4215080;  
RA Poljak R.J., Amzel L.M., Avey H.P., Chen B.L., Phisackerley R.P.,  
RA Saul F.;  
RT "The three-dimensional structure of the Fab' fragment of a human  
RL myeloma immunoglobulin at 2.0-A resolution";  
RL Proc. Natl. Acad. Sci. U.S.A. 71:3440-3444(1974).  
RN [8]  
RP PARTIAL SEQUENCE (BENCE-JONES PROTEIN MCG).  
RX MEDLINE=75013804; PubMed=4415202;  
RA Felt J.W., Deutesch H.F.;  
RT "Primary structure of the MCG lambda chain.";  
RL Biochemistry 13:4102-4114(1974).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF MCG.  
RA Edmundson A.B., Ely K.R., Abola E.E., Schiffer M.,  
RA Panagiotopoulos N.;  
RT "Rotational allotetramer and divergent evolution of domains in  
RL immunoglobulin light chains";  
RL Biochemistry 14:3953-3961(1975).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY OF MCG.  
RX MEDLINE=90133913; PubMed=2515285;  
RA Ely K.R., Herron J.N., Harker M., Edmundson A.B.;  
RT "Three-dimensional structure of a light chain dimer crystallized in  
RL water. Conformational flexibility of a molecule in two crystal  
RL forms.";  
RL J. Mol. Biol. 210:601-615(1989).  
RN [11]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=82080680; PubMed=6273747;  
RA Hieer P.A., Hollis G.F., Kormeyer S.J., Waldmann T.A., Leder P.;  
RT "Clustered arrangement of immunoglobulin lambda constant region genes  
RL in man.";  
RL Nature 294:536-540(1981).  
CC -1- MISCELLANEOUS: The sequence shown is the Kern-/Oz-/MCG- chain  
CC found in proteins SH, X, and NIG-64. The Kern protein has the  
CC Kern+ marker, the NEMM protein has the Oz+ marker, the MCG protein  
CC has the Kern+ marker, and the MCG+ marker.  
CC -1- MISCELLANEOUS: Six tandem lambda-type genes were identified and  
CC the 3 most 5' were sequenced. These correspond to the MCG sequence  
CC (lambda-1), the Kern-/Oz- sequence (lambda-2) and the Kern-/Oz+  
CC sequence (lambda-3).  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: J00253; AA59107.1; -  
DR EMBL: L28562; AAB36581.1; AAT\_INIT.  
DR EMBL: X51754; CAB38569.1; AAT\_INIT.  
DR EMBL: X51755; CAA36049.1; -  
DR EMBL: X51755; CAA36051.1; -  
DR PIR: A92057; L2HU.

DR PDB: 2MCG; 15-JUL-92.  
DR PDB: 7FAB; 31-JAN-94.  
DR PDB: 1AOK; 04-FEB-98.  
DR PDB: 1LIL; 15-MAR-97.  
DR Genew; HGNC:5855; IGLC1.  
DR Genew; HGNC:5856; IGLC2.  
DR Genew; HGNC:5857; IGLC3.  
DR MIM: 147220; -.  
DR GO; GO:0003823; F:antigen binding; NAS.  
DR GO; GO:0006935; P:Immune response; NAS.  
DR InterPro; IPR007110; IG-Like.  
DR InterPro; IPR003597; IG\_c1.  
DR InterPro; IPR003006; IG\_MHC.  
DR Pfam; PF00047; Ig\_c1; 1.  
DR SMART; SM00407; IGL1; 1.  
DR PROSITE; PS50835; IG\_LIKE; 1.  
DR PROSITE; PS00290; IG\_MHC; 1.  
KW Immunoglobulin domain; Immunoglobulin C region; Bence-Jones protein;  
KW 3D-structure.  
FT NON\_TER 1 1  
FT DOMAIN 6 100 IG-LIKE.  
FT DISULFID 27 86 INTERCHAIN (WITH HEAVY CHAIN).  
FT DISULFID 104 104 A-> N (IN MCG+ MARKER).  
FT VARIANT 5 5 /FTID=VAR\_003898.  
FT VARIANT 7 7 S-> T (IN MCG+ MARKER).  
FT VARIANT 45 45 S-> G (IN KERN+ MARKER).  
FT VARIANT 56 56 /FTID=VAR\_003900.  
FT VARIANT 56 56 T-> K (IN MCG+ MARKER).  
FT VARIANT 82 82 /FTID=VAR\_003901.  
FT VARIANT 82 82 R-> K (IN OZ+ MARKER).  
FT STRAND 8 11 /FTID=VAR\_003902.  
FT HELIX 15 19  
FT TURN 20 21  
FT STRAND 24 32  
FT STRAND 38 43  
FT TURN 44 45  
FT STRAND 46 48  
FT STRAND 52 59  
FT TURN 61 62  
FT STRAND 65 72  
FT HELIX 75 80  
FT STRAND 84 89  
FT STRAND 94 99  
SQ SEQUENCE 105 AA; 11236 MW; DCD9C7C201C13CC2 CRC64;  
Query Match 4.7%; Score 159; DB 1; Length 105;  
Best Local Similarity 34.1%; Pred. No. 0.0007;  
Matches 31; Conservative 18; Mismatches 40; Indels 2; Gaps 2;  
OY 521 OPR-EPQVYTPPPSDELTKQVSLTCLVKFYSDDIWEVNSGQP-ENNYKTPPVLYD 578  
DB 1 QPKAPSTVLRPPSEELQANKATIVCLISDFPEAPVAVKAKDSSPVKAGVETTPSKQ 60  
OY 579 SDGSFFLYSKLTVDKSRMKGQGVNPSGVNHE 609  
DB 61 SNKTYAASSYSLTPPEQWKSHRSYSCQVTHE 91  
RESULT 75  
LAC\_PIG STANDARD; PRT; 105 AA.  
AC P01846;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE IG lambda chain C region.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxId=9823;



```

RN [1]
RP MEDLINE=78000254; PubMed=409425;
RX Novotny J., Franek F., Margolies M.N., Haber E.;
RA "Amino acid sequence of normal (microheterogeneous) porcine
RT immunoglobulin lambda chains."
RL Biochemistry 16:3765-3772(1977).
CC -1- MISCELLANEOUS: This chain was obtained from a mixture of normal
CC immunoglobulins.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR: A02129; 11PG.
DR HSSP: P01842; 2MCG.
DR InterPro: IPR007110; IG-1-like.
DR InterPro: IPR003597; IG_C1.
DR InterPro: IPR003006; IG_MHC.
DR Pfam: PF00047; Ig_1.
DR SMART: SM00407; IG1; 1.
DR PROSITE: PS50835; IG_LIKE; 1.
DR PROSITE: PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT DOMAIN 1 100 IG-LIKE.
FT DISULFID 27 86
FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 105 AA; 11003 MW; 3817AAEBD747C396 CRC64;

Query Match 4.6%; Score 156; DB 1; Length 105;
Best Local Similarity 35.2%; Pred. No. 0.0011;
Matches 32; Conservative 15; Mismatches 42; Indels 2; Gaps 2;

OY 521 QPR-EPQYVTLPPSRDELTKNOVITCLVKGFPYSDIAVENESGQP-ENNYKTTTPVLD 578
DB 1 QPKAAPVTLPPSPSEELGTNKAIVLCISDFYPAAVTVWAKGTTVTCQVETTKSKQ 60
OY 579 SDGSFLYSLKLTVDKSRWQGNVFCSCVMHE 609
DB 61 SNKVAASSYIALSASDMKSSSGFTQVTHE 91

RESULT 76
KMLS_CHICK STANDARD; PRT; 1906 AA.
AC P11799; P19038;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin light chain kinase, smooth muscle and non-muscle isozymes
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP MEDLINE=96033976; PubMed=7589469;
RX Waterston D.W., Collings M., Lukes T.J., van Eldik L.J., Birukov K.G.,
RA Stepanova O.V., Shirinsky V.P.;
RT "Multiple gene products are produced from a novel protein kinase
RT transcritpion region."
RL FEBS Lett. 373:217-220(1995).
RN [2]
RP MEDLINE=90192792; PubMed=2315320;
RX Olson N.J., Pearson R.B., Needleman D.S., Hurwitz M.J., Kemp B.E.,
RA Means A.R.;
RT "Regulatory and structural motifs of chicken gizzard myosin light
RT chain kinase."
RL Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288(1990).
RN [3]
RP MEDLINE=90361738; PubMed=2202734;

```

```

RA Shoemaker M.O., Lau W., Shattuck R.L., Kwiatkowski A.P.,
RA Matrisian P.E., Guerra-Santos L., Wilson E., Lukes T.J.,
RA van Eldik L.J., Waterston D.W.;
RT "Use of DNA sequence and mutant analyses and antisense
RT oligodeoxynucleotides to examine the molecular basis of nonmuscle
RT myosin light chain kinase autoinhibition, calmodulin recognition, and
RT activity."
RL J. Cell Biol. 111:1107-1125(1990).
RN [4]
RP MEDLINE=92236611; PubMed=1373815;
RX Collinge M., Matrisian P.E., Zimmer W.E., Shattuck R.L., Lukes T.J.,
RA van Eldik L.J., Waterston D.W.;
RT "Structure and expression of a calcium-binding protein gene contained
RT within a calmodulin-regulated protein kinase gene."
RL Mol. Cell. Biol. 12:2359-2371(1992).
CC -1- FUNCTION: Phosphorylates a specific serine in the N-terminus of a
CC myosin light chain, which leads to the formation calmodulin/MLCK
CC signal transduction complexes which allow selective transduction
CC of calcium signals.
CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
CC light-chain] phosphate.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative Initiation;
CC Comment=At least 3 isoforms, MLCK-210/Non-muscle,
CC MLCK-108/Smooth-muscle and Telokin, are produced by alternative
CC initiation;
CC -1- TISSUE SPECIFICITY: ISOFORM TELOKIN IS EXPRESSED IN GIZZARD,
CC HEART, LUNG, INTESTINE, AND SKELETAL MUSCLE ALTHOUGH THE LEVELS OF
CC THE EXPRESSION IN THE LATTER WERE MUCH LESS THAN THAT IN THE
CC GIZZARD.
CC -1- DOMAIN: TELOKIN BINDS CALMODULIN.
CC -1- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC -1- SIMILARITY: Contains 9 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52876; CA37056.1; -
DR EMBL: X52876; CA37057.1; -
DR EMBL: X52876; CA37058.1; -
DR EMBL: M31048; AAA49069.1; -
DR EMBL: M14953; AAA69964.1; -
DR EMBL: M96655; AAA48083.1; -
DR EMBL: M88283; AAA48647.1; -
DR EMBL: M88284; AAB53768.1; -
DR PIR: S68235; S68235.
DR PDB: 1CDL; 31-ANG-24.
DR PDB: 1YRK; 27-APR-99.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR007110; IG_1-like.

```



[illegible]

QY	89	QGNFPLIKNIKLIEDSDTYICEVED-----QKEEYVOLLVFGILASDSHLLQGGSLTLE	144
Db	81	RGFSFLVITGVGVGDSGSKYTCEAANDGVRQVLYVELVVEG-----NS-----LKYSLPSAK	133
QY	145	SPGQ--SSPSVQGR-----SPRGNIQGGKT-----	168
Db	134	TPGGRSLSVPRVHRPSPINGESPSPKATKPNRVVAREGGTGAFSCKITGRPQPVYTWTKGD	193
QY	169	-----LSVSQLELQDSGTWTCTVLQNOKVVEFKIDIVY-----	201
Db	194	IHLQONERFNMFEKGTGIQYLEIQNVQLADAGIYCTFTVNSAGKASVAELTVQGPDKTDT	253
QY	202	-----LA-----FOKASIIYKKKEGEQVESFPLATVETLTCSSGELMW	240
Db	254	HAQPLCPMPKPTTLAKVLAENSDFOATNSNGIAELKSTHEL-MVEFKRLSKKKEFTY	312
QY	241	QAEASASSKSWITFDLKNKVEVSKVNTODPKLQMKCKPLHLTLPOALPOVYASGNLTLLA	300
Db	313	TSEKADGQGGQNGQENANVPLQESGKGTGPQLQKTSITLQAVNKAQPEKKAEPQTTFI	372
QY	301	LEAKTGKLGHEVNLVVMRATQLOKNLTCGEWGPSTPKMLSLKLEN-----KEAKYSKRE	355
Db	373	RQAEKDRK--RTVQPLMTTTTQGENPSLT-----QVQSPR--SRETFENAGVRAKSVKEBRE	423
QY	356	KPVWVNLVPEKGMQCLLSDSGQ-VLLESNIKLP---TW---STVPEPKSC-----DK	401
Db	424	PLGIPRPFESRPOSLESEGGQEIFKSKVSGKRPDVEWMEKEGVPIKTGEGIQIYEBDG	482
QY	402	THTC-----PPCAPELLGSPSVLPFPKPKDT---LMSIRPEVT-----CVVV	443
Db	483	THCLMLKKACLGDS--GSYSCAAFPRPGQTSISMLITVAKRKVEVAPCRSSVLKGTIV-	539
QY	444	DVSHEDPEYKFMNYVDGVENNAKTRPRE--QY-NSTVRV-VSVLTVLHDQWLNGKE--	497
Db	540	---SEGQDFLQYGVGVVPRPEITWMLNEQPIQVASHSTFEAGVAKLTV--QDALPEDDGI	594
QY	498	YKCKVSNKA-----LPALEKITSKAKGQPREPVYTLTP-----SRDELTKQV	542
Db	595	YTCLAEINMAGRASCSAQVTVKEKSKSKAEKGQAANKTKTAPAFILKGLTDLTKYMDSQV	654
QY	543	SLTCLVKGFPSPDIAYEWESNQPE-----NNYKTPPLTD-----SDGSFFLYSKLTV	591
Db	655	IMT-----VESANPCPEIIMLHNGKEIQETEDPHFEKKGNEYSLYIQEVPE	702
QY	592	DKSRWQGGNVFSCSVNHEALHNHYTOKSLSPGLQDLETCAEADQD	638
Db	703	DTGK-----YTCEANNE-LGSETQIQATL-----TVQEPRQD	732
RESULT 77			
DTC_HUMAN			
ID	DTC_HUMAN	STANDARD;	PRT; 383 AA.
AC	P01880;		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	21-JUL-1986 (Rel. 01, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	IG delta chain C region.		
GN	IGHD.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE OF 1-162 (MYELOMA PROTEIN WAH).		
RX	MEDLINE=82082419; PubMed=6947220;		
RA	Putnam F.W., Takahashi N., Teter D., Debuire B., Lin L.-C.;		
RT	"Amino acid sequence of the first constant region domain and the		
RT	hinge region of the delta heavy chain of human IgD."		
RL	Proc. Natl. Acad. Sci. U.S.A. 78:6168-6172(1981).		
RN	[2]		
RP	SEQUENCE OF 158-363 (MYELOMA PROTEIN WAH).		
RX	MEDLINE=81237368; PubMed=6787589;		
RA	Lin L.-C., Putnam F.W.;		



```

RT "Primary structure of the Fc region of human immunoglobulin D:
RT implications for evolutionary origin and biological function.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:504-508(1981).
RN [3]
RP SEQUENCE OF 103-137 (MYELOMA PROTEIN NIG-65), AND CARBOHYDRATES.
RX MEDLINE=82231401; PubMed=7092891;
RA Takayasu T., Suzuki S., Kametani F., Takahashi N., Shinoda T.,
RA Okuyama T., Munekata E.;
RT "Amino acid sequence of galactosamine-containing glycopeptides in the
RT hinge region of a human immunoglobulin D.";
RL Biochem. Biophys. Res. Commun. 105:1066-1071(1982).
RN [4]
RP SEQUENCE OF 158-383 (MYELOMA PROTEIN NIG-65).
RX MEDLINE=81199406; PubMed=6785754;
RA Shinoda T., Takahashi N., Takayasu T., Okuyama T., Shimizu A.;
RT "Complete amino acid sequence of the Fc region of a human delta
RT chain.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:785-789(1981).
RN [5]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=8113632; PubMed=7008791;
RA Takayasu T., Takahashi N., Shinoda T.;
RT "Amino acid sequence and location of the three glycopeptides in the
RT Fc region of human immunoglobulin D.";
RL Biochem. Biophys. Res. Commun. 97:635-641(1980).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
DR PIR, A02175; DHU.
DR Genes: HGNC:5480; IGHD.
DR MIM, 147170;
DR GO, GO:0003823; F:antigen binding; NAS.
DR GO, GO:0006955; P:immune response; NAS.
DR InterPro: IPR007110; IG-1-like.
DR InterPro: IPR003597; IG-cl.
DR InterPro: IPR003006; IG_MHC.
DR Pfam, PF00047; Ig_3.
DR SMART, SM00407; IGcl; 3.
DR PROSITE, PS50835; IG_LIKE; 3.
DR PROSITE, PS00290; IG_MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT NON TER 1 1
FT DOMAIN 6 98
FT 175 263 IG-LIKE 1.
FT DOMAIN 267 373 IG-LIKE 2.
FT DISULFID 15 15 IG-LIKE 3.
FT DISULFID 28 84 INTERCHAIN (WITH LIGHT CHAIN).
FT CARBOHYD 109 161 INTERCHAIN (WITH HEAVY CHAIN).
FT CARBOHYD 110 109 O-LINKED.
FT CARBOHYD 110 110 O-LINKED.
FT CARBOHYD 113 113 O-LINKED.
FT CARBOHYD 126 126 O-LINKED.
FT CARBOHYD 127 127 O-LINKED.
FT CARBOHYD 131 131 O-LINKED.
FT CARBOHYD 132 132 O-LINKED (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .).
SQ SEQUENCE 383 AA; 42125 MW; DAIDET3519CT6CID CRC64;
Query Match 4.5%; Score 154; DB 1; Length 383;
Best Local Similarity 23.2%; Pred. No. 0.008; Indels 82; Gaps 11;
Matches 63; Conservative 34; Mismatches 92;

```

```

OY 527 ----VYTLPPSRDEL-TRKQVSLTCLVKGFPSDIAVEMESNGQ-PENNYKTPPVLDSD 580
DB 166 QPLGVILTPAVQDLMARDKATFTCFVVGSDLDMDAHLTWAVAGVPTGVBEGLEHNSN 225
OY 581 GSFFLYSKLTVDKSRWQGVFSCSVMEAL 611
DB 226 GSQGSRLTLPRSLMAGTSVCTLHNPSTL 256
RESULT 78
LACS_MUASP STANDARD; PRT; 105 AA.
AC P20765;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda-5 chain C region.
OS Mus spretus (western wild mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88196070; PubMed=3129289;
RA Mani F., Cazenave P.A., Kindt T.J.;
RT "Conservation of the immunoglobulin C lambda 5 gene in the Mus gene.";
RL EMBO J. 7:117-122(1988).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M35582; AAA39152.1; -.
DR HSSP, P01842; 2MCG.
DR InterPro: IPR007110; Ig-1-like.
DR InterPro: IPR003597; IG-cl.
DR InterPro: IPR003006; IG_MHC.
DR Pfam, PF00047; Ig_1.
DR SMART, SM00407; IGcl; 1.
DR PROSITE, PS50835; IG_LIKE; 1.
DR PROSITE, PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1 1
FT DOMAIN 6 100
FT DISULFID 27 86 IG-LIKE.
FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 105 AA; 11674 MW; AAB417DF68471A17 CRC64;
Query Match 4.5%; Score 153.5; DB 1; Length 105;
Best Local Similarity 34.0%; Pred. No. 0.0016;
Matches 35; Conservative 20; Mismatches 43; Indels 5; Gaps 3;

```



```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE IG_Lambda chain C region.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OK NCBI_TaxId=9986;
RN (1)
RX MEDLINE=82091105; PubMed=6797414;
RA Garcia I., Jacon J.-C.;
RT "The primary structure of the constant region of Basilea-rabbit
RL Immunoglobulin lambda-chains."
BL Biochem. J. 197;177-183 (1981).
CC -1- MISCELLANEOUS: This lambda chain expresses the c7 allotypic
CC specificity.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02130; L7RB.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; FALSE NEG.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 6 100 IG-LIKE.
FT DISULFID 27 86
FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 105 AA; 11484 MW; B427513272E8663D CRC64;

```

```

Query Match 4.5%; Score 153; DB 1; Length 105;
Best Local Similarity 32.7%; Pred. No. 0.0017;
Matches 32; Conservative 17; Mismatches 45; Indels 4; Gaps 2;

```

```

QY 525 POVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP-ENNYKTPPLVDSGSP 583
DB 6 PEVILFPSPSEELKDKKATLVCLISDFYRIVKVMKADGNSVTQGVDTTQSKSNNKY 65
QY 584 FLYSKLTVDKSRMQGNSVSCVMHEALHNHYTKSL 621
DB 66 AASFHLTANQWKSYQSVTCQVTHE---GHTVEKSLA 100

```

```

RESULT 80
KCB RABIT STANDARD; PRT; 106 AA.
AC P01839;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE IG_kappa-B4 chain C region.
GN K-BAS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OK NCBI_TaxId=9986;
RN (1)
RX MEDLINE=82091105; PubMed=6797414;
RA Garcia I., Jacon J.-C.;
RT "The primary structure of the constant region of Basilea-rabbit
RL Immunoglobulin lambda-chains."
BL Biochem. J. 197;177-183 (1981).
CC -1- MISCELLANEOUS: In Basilea rabbits, the major type of light chain
CC is lambda. The kappa chain shown is a minor component. All other
CC rabbit B allotypes have Cys-64
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

```

DR EMBL; V01241; CAA24558.1;
DR EMBL; V00885; -; NOT_ANNOTATED_CDS.
DR PIR; A02121; K4RBS.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; FALSE NEG.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 6 99 IG-LIKE.
FT DISULFID 27 87
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
SQ SEQUENCE 106 AA; 11279 MW; AF9B928DDA853849 CRC64;

```

```

Query Match 4.5%; Score 153; DB 1; Length 106;
Best Local Similarity 33.7%; Pred. No. 0.0017;
Matches 31; Conservative 19; Mismatches 36; Indels 6; Gaps 2;

```

```

QY 520 GQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP-ENNYKTPP 575
DB 1 GDPVAPSVLTPPPSKELITGTATIVCANRFPYSDITVTKVKGTQSGIENSKT--P 58
QY 576 VLDSDGPFYLSKLTVDKSRMQGNSVSCVM 607
DB 59 QSPEDNTYLSSTJLSIAQYNHSHSVTCVAV 90

```

```

RESULT 81
VCAM1 MOUSE STANDARD; PRT; 739 AA.
ID VCAM1 MOUSE
AC P29533;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vascular cell adhesion protein 1 precursor (V-CAM 1).
GN VCAM1 OR VCAM-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OK NCBI_TaxId=10090;
RN (1)
RX MEDLINE=92181437; PubMed=1371918;
RX Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wyek M.,
RA Burdly L., Miyake K., Kincaide P., Lobb R.;
RT "Cloning of murine and rat vascular cell adhesion molecule-1."
RL Biochem. Biophys. Res. Commun. 183:163-169 (1992).
RN (2)
RX MEDLINE=93246254; PubMed=7683304;
RX Araki M., Araki K., Vasealli P.;
RT "Cloning and sequencing of mouse VCAM-1 cDNA."
RL Gene 126:261-264 (1993).
RN (3)
RX MEDLINE=94117008; PubMed=7507076;
RX Cybulsky M.I., Allan-McCormack M., Collins T.;
RT "Structure of the murine VCAM1 gene."
RL Genomics 18:387-391 (1993).
RN (4)
SQ SEQUENCE OF 1-693 FROM N.A. (ISOFORM 1).

```











RP PHOSPHORYLATION OF TYR-477 AND TYR-501, AND MUTAGENESIS OF TYR-436;  
 RYR-460; TYR-477 AND TYR-501.  
 RX MEDLINE=98204923; PubMed=9535915;  
 RA Takeda T., Macozaki T., Takeda H., Fukunaga K., Noguchi T.,  
 RA Takeda Y., Okazaki I., Tsuda M., Yamao T., Ochi F., Kasuga M.,  
 RT "Roles of the complex formation of SHP-1 with SHP-2 in  
 RT insulin-stimulated mitogen-activated protein kinase activation.";  
 RL J. Biol. Chem. 273:9234-9242 (1998).  
 CC -1- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts  
 CC as docking protein and induces translocation of PRPN6,  
 CC PTEN11 and other binding partners from the cytosol to the  
 CC plasma membrane. Supports adhesion of cerebellar neurons, neurite  
 CC outgrowth and glial cell attachment. May play a key role in  
 CC intracellular signaling during synaptogenesis and in synaptic  
 CC function. Involved in the negative regulation of receptor tyrosine  
 CC kinase-coupled cellular responses induced by cell adhesion, growth  
 CC factors or insulin. Mediates negative regulation of phagocytosis,  
 CC mast cell activation and dendritic cell activation. CD47 binding  
 CC prevents maturation of immature dendritic cells and inhibits  
 CC cytokine production by mature dendritic cells. May play a role in  
 CC the release of nitric oxide by macrophages (By similarity).  
 CC -1- SUBUNIT: Binds PRPN11 when tyrosine-phosphorylated, except in  
 CC macrophages, where it primarily binds PRPN6. Binds GBB2 in vitro.  
 CC Binds FGR. Binds JAK2 irrespective of its phosphorylation status  
 CC and forms a stable complex. Binds SCAP1 and/or SCAP2. The  
 CC resulting complex recruits Fyb. Binds PRK2B (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in brain, spleen, lung, liver  
 CC and kidney. Detected at lower levels in heart. Highly expressed in  
 CC alveolar and peritoneal macrophages, and at lower levels in  
 CC dendritic cells.  
 CC -1- PTM: N-glycosylated.  
 CC -1- PTM: Phosphorylated on tyrosine residues in response to insulin,  
 CC cell adhesion or epidermal growth factors. Dephosphorylated by  
 CC PRPN11.  
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D85183; BAA12734.1; -;  
 DR EMBL: D38468; BAA20368.1; -;  
 DR EMBL: U62328; AAC68478.1; -;  
 DR EMBL: AF055065; AAC18089.1; -;  
 DR HSSP: P01703; 7PAB.  
 DR InterPro: IPR007110; IG-1-like.  
 DR InterPro: IPR003597; IG\_C1.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR Pfam: PF00047; Ig\_3.  
 DR SMART: SM00407; IG1; 2.  
 DR PROSITE: PS50835; IG\_LIKE; 3.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 KW Repeat: Signal; Transmembrane; Immunoglobulin domain; SH2-binding;  
 KW Glycoprotein; Phosphorylation.  
 FT SIGNAL 1 31  
 FT CHAIN 32 509  
 FT DOMAIN 28 373  
 FT TRANSMEM 374 394  
 FT DOMAIN 395 509  
 FT DOMAIN 32 138  
 FT DOMAIN 150 248  
 FT DOMAIN 255 349  
 FT DISULFID 55 122  
 FT DISULFID 172 229  
 FT DISULFID 274 332  
 FT SITE 436 439  
 FT SH2-BINDING (POTENTIAL).

FT SITE 446 451 SH2-BINDING (POTENTIAL).  
 FT SITE 460 463 SH2-BINDING (POTENTIAL).  
 FT SITE 477 480 SH2-BINDING (POTENTIAL).  
 FT SITE 501 504 SH2-BINDING (POTENTIAL).  
 FT MOD\_RES 436 436 PHOSPHORYLATION (BY TYR-KINASES)  
 FT MOD\_RES 460 460 PHOSPHORYLATION (BY TYR-KINASES)  
 FT MOD\_RES 477 477 PHOSPHORYLATION (BY TYR-KINASES).  
 FT MOD\_RES 501 501 PHOSPHORYLATION (BY TYR-KINASES).  
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT MUTAGEN 436 436 Y-P: ABOLISHES TYROSINE PHOSPHORYLATION  
 FT AND PTEN11 BINDING; WHEN ASSOCIATED WITH  
 FT F-460, F-477 AND F-501.  
 FT MUTAGEN 460 460 Y-P: ABOLISHES TYROSINE PHOSPHORYLATION  
 FT AND PTEN11 BINDING; WHEN ASSOCIATED WITH  
 FT F-436, F-477 AND F-501.  
 FT MUTAGEN 477 477 Y-P: STRONGLY REDUCES INSULIN-INDUCED  
 FT TYROSINE PHOSPHORYLATION AND PTEN11  
 FT BINDING. ABOLISHES TYROSINE  
 FT PHOSPHORYLATION AND PTEN11 BINDING; WHEN  
 FT ASSOCIATED WITH F-436, F-460 AND F-501.  
 FT Y-P: STRONGLY REDUCES INSULIN-INDUCED  
 FT TYROSINE PHOSPHORYLATION AND PTEN11  
 FT BINDING. ABOLISHES TYROSINE  
 FT PHOSPHORYLATION AND PTEN11 BINDING; WHEN  
 FT ASSOCIATED WITH F-436, F-460 AND F-477.  
 FT P -> L (IN REF. 4).  
 FT MISSING (IN REF. 4).  
 FT F -> I (IN REF. 3).  
 FT S -> C (IN REF. 4).  
 FT KR -> MP (IN REF. 3; AA SEQUENCE).  
 FT D -> A (IN REF. 2).  
 FT D -> N (IN REF. 3).  
 FT N -> L (IN REF. 3; AA SEQUENCE).  
 FT N -> G (IN REF. 3; AA SEQUENCE).  
 FT G -> F (IN REF. 3; AA SEQUENCE).  
 FT E -> P (IN REF. 3; AA SEQUENCE).  
 FT NARE -> EGON (IN REF. 3; AA SEQUENCE).  
 FT R -> E (IN REF. 3; AA SEQUENCE).  
 FT MISSING (IN REF. 3; AA SEQUENCE).  
 FT SEQUENCE 509 AA; 55690 MM; 5BE1PE0A4D0429F4 CRC64;  
 SQ  
 Query Match 4.4%; Score 150; DB 1; Length 509;  
 Best Local Similarity 21.6%; Pred. No. 0.021;  
 Matches 83; Conservative 63; Mismatches 158; Indels 80; Gaps 16;  
 Oy 259 KEYSVKRVTDPPKIQNGKKLPLHLTLFQALP-----OYAGSGNLTLLAKTGKLAQEVN 313  
 Db 32 KELVVTQADSDVSVAAGDSATLNCVTSSLPVGPRIKFKGEGGNRSPISYFIDGEHPRI 91  
 Oy 314 LVVWATQLOKNTLCEVWGTPSKMLSLTKENKEAVSRKRPVWVLANPAGMOCULS 373  
 Db 92 TWSVDAIK-RNND-----FSICISN-----VTPEDAGIYYVKF 125  
 Oy 374 DSGGVLLSNIK-----VLPTWSTPVEPKSCDKHTTQPCPAPRLGSPVFLPPK 425  
 Db 126 QKGIVEPDTIKSGGGTLLV-----AKPSS-----PEVSGPSRSGP----- 164  
 Oy 426 PKOTLMISRPPEVTCVVVDVSHEDPEVKFWMYVDGVEVHNAKTKRREQDSTYRVSVL 485



```

Db 165 -----GQTVNFTCKSYGFSPPN--ITLKLKDGKELSHLETTISKS-NVSYNISTV 214
Qy 486 TV-LHODMNGKXKCVSNKALPA-PIEKTISKAGQPREPOVYTLPPSRDEL-T-KNOV 542
Db 215 SYKLSEEDHSKRI-CEVAHVTLGKPLNGTANFSIIVSP---TLTKTQOPLTPASOV 270
Qy 543 SLTCLVKGFPSPDIAVWESNGCPENNYKTPPVLDSDGSFFLYSKLTVYDKSRMOGNYF 602
Db 271 NLTCQVQKFPKALQNLNLENGLSRTDKPEHFTDRDGTYNVYTLFLVNSSAHREDYVF 330
Qy 603 SCGVNHE---ALNHTYTKSLSLIS 623
Db 331 TCVEHDSQPAITENTHTVAFNHS 354

RESULT 84
HB2C_PIG STANDARD; PRT; 261 AA.
ID HB2C_PIG
AC P15982;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SLA class II histocompatibility antigen, DQ haplotype C beta chain
DE precursor.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=90361905; PubMed=2391424;
RA Gustafsson K., Leguern C., Hirsch F., Germana S., Pratt K.,
RA Sachs D.H.;
RT "Class II genes of miniature swine. IV. Characterization and
RT expression of two allelic class II DQB cDNA clones.";
RL J. Immunol. 145:1946-1951(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M31497; AAA31084.1; -
DR EMBL: M32117; AAA53110.1; -
DR PIR: A60404; A60404.
DR HSSP: P13760; 2SEB.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003597; IG-cl.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR000353; MHC_II_beta.
DR Pfam: PF00047; IG_1.
DR Pfam: PF00969; MHC_II_beta; 1.
DR ProDom: PD000328; MHC_II_beta; 1.
DR SMART: SM00407; IGc1_1.
DR PROSITE: PSS0835; IG_LIKE; 1.
DR PROSITE: PS00290; IG_MHC; 1.
DR MHC II; Transmembrane; Glycoprotein; Signal.
KW SIGNAL.
FT CHAIN 1 31
FT 32 261 SLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 32 126 DQ HAPLOTYPE C BETA CHAIN.
FT DOMAIN 127 220 EXTRACELLULAR BETA-1.
FT DOMAIN 221 230 EXTRACELLULAR BETA-2.
FT TRANSMEM 231 251 CONNECTING PEPTIDE.
FT DOMAIN 252 261
FT DISULFID 47 111 CYTOPLASMIC TAIL.
FT DISULFID 149 205 BY SIMILARITY.
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 261 AA; 29550 MW; AA9581F2A3B1969D CRC64;

```

```

Query Match 4.4%; Score 149; DB 1; Length 261;
Best Local Similarity 24.8%; Pred. No. 0.01;
Matches 51; Conservative 36; Mismatches 99; Indels 20; Gaps 6;

Qy 409 PAPELLGSPVFLPPPKKDTLMSRPEVTCVYVVDVSHDEPEKFNMYVDGVEYHNAKT 468
Db 28 PVAERDRSPDFVFOFK-GECEFYNGTQVRGVARYIYNQEHILRFDSDVEFRAVTPLG 86
Qy 469 KPREQVSTTRVYVVLVTLVHODMNGKXKCVSNKALPAPIEKTISKAGQPREPOVY 528
Db 87 READSWNSQKDVLEQMPA-EVDRCVKNYQLE-----ESTTLQRRVQ---TV 131
Qy 529 TLPPSRDELITKQVSLTCLVKGFPSPDIAVWESNGCPEN-NYKTPPVLDSDGSFFLYS 587
Db 132 TISPESKALALHNLNLVCAVVDIFYPSQVQVQWFRNGQETGAGVSTPLIRNGD---WTYQ 188
Qy 588 KLTVDKSRMOGNYFSCGVNHEALHN 613
Db 189 VLVMLENNLORGDDVYTCRVEHSLQN 214

RESULT 85
DSCA_HUMAN STANDARD; PRT; 2012 AA.
ID DSCA_HUMAN
AC O60469; O60468;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Down syndrome cell adhesion molecule precursor (CHD2).
DE DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=98087574; PubMed=9426258;
RA Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
RA Lyons G.E., Korenberg J.R.;
RT "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT Down syndrome region and is involved in the development of the
RT nervous system.";
RL Hum. Mol. Genet. 7:227-237(1998).
RN (2)
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE=20384934; PubMed=10925149;
RA Agarwala K.L., Nakamura S., Teutsuni Y., Yamakawa K.;
RT "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT intercellular adhesion.";
RL Brain Res. Mol. Brain Res. 79:118-126(2000).
RN (3)
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sakaki T., Nagamine K., Mitsuyma S., Antonarakis S.E.,
RA Mitsuhashi S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecher H.,
RA Ramer J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
CC -I- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC NERVOUS SYSTEM DEVELOPMENT.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC SHORT ISOFORM MAY BE SECRETED.

```



```

CC -1- ALTERNATIVE PRODUCTS;
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long; Synonyms=CHD2-52;
CC IsoId=O60469-1; Sequence=Displayed;
CC Name=Short; Synonyms=CHD2-42;
CC IsoId=O60469-2; Sequence=VSP_002502, VSP_002503;
CC -1- TISSUE SPECIFICITY: Primarily expressed in brain.
CC -1- SIMILARITY: Contains 10 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 6 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF023450; AAC17967.1; -
DR EMBL; AF023449; AAC17966.1; -
DR EMBL; AF217525; AAF27525.1; -
DR EMBL; AL163283; CAB90464.1; -
DR EMBL; AL163282; CAB90436.1; -
DR EMBL; AL163281; CAB90444.1; -
DR GeneW; HGNC:3039; DSCAM.
DR MIM; 602523; -
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0007155; P:cell adhesion; TAS.
DR GO; GO:0007399; P:neurogenesis; TAS.
DR InterPro; IPR003957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG_1-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00041; fn3; 6.
DR Pfam; PF00047; igf; 9.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00408; IGC2; 7.
DR PROSITE; PS00835; IG_LIKE; 9.
DR Immunoglobulin domain; Glycoprotein; signal; Cell adhesion; Repeat;
KW Transmembrane; Alternative splicing.
FT SIGNAL 1 17
FT CHAIN 18 2012
FT DOMAIN 18 1595
FT TRANSMEM 1596 1616
FT DOMAIN 1617 2012
FT DOMAIN 39 129
FT DOMAIN 125 216
FT DOMAIN 225 305
FT DOMAIN 313 401
FT DOMAIN 407 500
FT DOMAIN 504 592
FT DOMAIN 596 685
FT DOMAIN 690 783
FT DOMAIN 787 883
FT DOMAIN 885 972
FT DOMAIN 984 1076
FT DOMAIN 1088 1177
FT DOMAIN 1189 1273
FT DOMAIN 1285 1377
FT DOMAIN 1380 1463
FT DOMAIN 1477 1562
FT DISULFID 46 102
FT DISULFID 145 197
FT DISULFID 246 293
FT DISULFID 335 385
FT DISULFID 428 484
FT DISULFID 525 575
FT DISULFID 617 669
FT DISULFID 711 766
FT DISULFID 809 865
FT DISULFID 1307 1359
FT CARBOHYD 28 28

```

```

FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 487 487 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 512 512 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 658 658 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 795 795 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1142 1142 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1160 1160 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1250 1250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1271 1271 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1468 1468 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1562 1571 N-GLYCOSYLATION (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 1572 2012 /FtId=VSP_002502.
FT VARSPLIC 1893 2012 Missing (in isoform Short).
FT CONFLICT 1893 2012 /FtId=VSP_002503.
FT CONFLICT 1893 2012 HRRDRLHLPYILMDPLNRGGRGTSRDLGACLEPOK
FT FT SRTLKRPVLEPIEMEAASSSTRGOSQWPGAVLPOR
FT FT EGAEIGQAAKMSQSESLDSRGLKGNPNYASVTLV ->
FT FT IGVTSYICHTLEWFC (IN REF. 1).
FT SO SEQUENCE 2012 AA; 222259 MW; 0E33CFB781A08334 CRC64;

Query Match 4.3%; Score 148.5; DB 1; Length 2012;
Best Local Similarity 20.3%; Pred. No. 0.16;
Matches 130; Conservative 78; Mismatches 230; Indels 203; Gaps 32;

34 GDTVELCTASQKSIQFHKNSNQIKILNQSGFLTKGSKLNDRAISRRLDQGNFP 93
239 GQRYELCKLGHPEPRYLKDNM-----PLELSGR-----FQKVTYG 277
94 LIINKLKIEDSDTYICEVDQKEVOLVFGLTANSPTHLLOQSLTTLT-----ESPFG 148
278 LIENINRPSDGSVCVSNR-----YG-TAKVIGRLVYVQPLKATISPRKVSVG 328
149 SPSVQCR-----SPRKNIQ-----GKITSVSLBLQDSGTWTCV 186
329 SQVSLSCVSTGTEDELSWRNGEILNP-GKNVITGINHEMLIMDMVKSDDGAYOCFV 387
187 LONQKVEFKIDIVLAFQKASSIVYKKEGOVFSFPLFVTEKLTGSGELMWQARAS 246
388 RKDLSAQDYVQVL-----EDTPKILS-AF-----SKRV 418
247 SSKSWITFDLKNKEVSKVLT---ODPKLQMK-----KLPLHLTLPOALPOY 291
419 SPAEPVSLMCNVKGTPLPTITWLLDDPILKGSISRISQMTSEGNVSYLNISSQVDR 478
292 AGSNLTLALAEATGKLHGEVNLVWRKATOLQKLT-----CEV-----NG 332
479 GGVYRCYANNASAGVLYQARINVRGPASIRPMKNITIAIGRDYYIHGRVIGYPYSIKWY 538
333 PTPSKMLSLKL-----ENKEAKVSKKEKPYVNLNPEAGMOCCLSDSGOYLIESIKVLP 388
539 KNSULPLPFRNRQVAFENKGLKLSDVQKEY-----DEGEVTC-----NLVQOP 581
389 TWSTVPVPSKCDKTHTCPPCAPELLGSPVFLFPPPKDXTLWISRTPEVTCVVVDVSH 448
582 QLSLT---SQSVHTVVKVPPFIQ-----FEPP-----RFSIGQVFIQCVVVS---G 622
449 DPEVKFMYVD-----GVEVNAKTKREQGYNSTRVSVLYTLHODMLNGKRYKC 500
623 DLPTITWQKDGPRIPESLGVITDNI-----DFTSSLR-ISLSTLMH-----NG-NYTC 669
501 KVSNAKLPARIEKTISSAKGQPREPOVYTLPSRDELTKNQVSLTCLVKGPYPSDIAVEW 560
670 IARNEA--AAVEHQSLIVAVP--PKRVVQPRQDGIYGAAVILNCSAEG-YFVPTIWMK 724
561 ESNQ-----QP-----ENNYKTPPVLDSDGSPFL 585

```



Db 725 FSKGAGVQFQFIALNGRIQVLNSGLIKHVEEDSGYLL 765

```
RESULT 86
KAC_HUMAN
ID KAC_HUMAN STANDARD; PRT; 106 AA.
AC P01834;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 1g kappa chain C region.
GN IGKC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCB1_TaxId9606;
RN 11)
RP SEQUENCE (MYELOMA PROTEIN EU).
RX MEDLINE=71064023; PubMed=5489770;
RA Gottlieb P.D., Cunningham B.A., Rutishauser U., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. VI. Amino
RT acid sequence of the light chain.";
RL Biochemistry 9:3155-3161(1970).
RN 12)
RP DISULFIDE BONDS.
RX MEDLINE=71064027; PubMed=4923144;
RA Call W.E., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. X.
RT intrachain disulfide bonds.";
RL Biochemistry 9:3188-3196(1970).
RN 13)
RP SEQUENCE (BENCE-JONES PROTEIN TI).
RX MEDLINE=72188439; PubMed=50277703;
RA Suter L., Barnikol H.U., Watanabe S., Hilschmann N.;
RT "Rule of antibody structure. The primary structure of a monoclonal
RT gamma globulin L-chain of kappa-type, subgroup 3 (Bence-Jones protein
RT TI). IV. The complete amino acid sequence and its significance for
RT the mechanism of antibody production.";
RL Hoppe-Seyler's Z. Physiol. Chem. 353:189-208(1972).
RN 14)
RP SEQUENCE FROM N.A.
RX MEDLINE=81042304; PubMed=6775818;
RA Hleber P.A., Max E.E., Seidman J.G., Maizel J.V. Jr., Leder P.;
RT "Cloned human and mouse kappa immunoglobulin constant and J region
RT gene conserve homology in functional segments.";
RL Cell 22:197-207(1980).
RN 15)
RP SEQUENCE (BENCE-JONES PROTEIN ROY).
RA Hilschmann N., Barnikol H.U., Hess M., Langer B., Ponertingl H.,
RA Steinhilber-Kayne M., Suter L., Watanabe S.;
RL (in French F., Stugart D. (eds.))
RL Gamma globulins: structure and function, pp.57-74, Academic Press,
RL New York (1969).
RN 16)
RP SEQUENCE (BENCE-JONES PROTEIN CUM).
RX MEDLINE=68242259; PubMed=5586923;
RA Hilschmann N.;
RT "The complete amino acid sequence of Bence Jones protein Cum (kappa-
RT type)."
RT Hoppe-Seyler's Z. Physiol. Chem. 348:1718-1722(1967).
RN 17)
RP SEQUENCE (BENCE-JONES PROTEIN AG).
RX MEDLINE=69234734; PubMed=4893682;
RA Titani K., Shinoda T., Putnam F.W.;
RT "The amino acid sequence of a kappa type Bence-Jones protein. 3. The
RT complete sequence and the location of the disulfide bridges."
RL J. Biol. Chem. 244:3550-3560(1969).
RN 18)
RP SEQUENCE (WALDENSTROM'S MACROGLOBULIN OU).
RX MEDLINE=70201507; PubMed=5447531;
RA Kohler H., Shimizu A., Paul C., Putnam F.W.;
RT "Macroglobulin structure: variable sequence of light and heavy
```

```
RT chains."
RL Science 169:56-59(1970).
RN 19)
RP SEQUENCE OF 1-33, 38-41 AND 62-80.
RC TISSUE=Abdominal adipose tissue;
RX MEDLINE=98249779; PubMed=9588180;
RA Olsen K.E., Sletten K., Westermarck P.;
RT "Extended analysis of AL-amyloid protein from abdominal wall
RT subcutaneous fat biopsy: kappa IV immunoglobulin light chain."
RL Biochem. Biophys. Res. Commun. 245:713-716(1998).
CC -I- MISCELLANEOUS: The EU sequence has the INV (3) allotypic marker,
CC 45-Ala and 83-Val. The ROY sequence has the INV (1,2) allotypic
CC marker, 45-Ala and 83-Leu.
CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00241; AAA58989.1; -.
DR EMBL; V00557; CAA23823.1; -.
DR PIR; B90562; K3HU.
DR PDB; 1D5B; 09-FEB-00.
DR PDB; 1D5I; 09-FEB-00.
DR PDB; 1D6V; 04-OCT-00.
DR PDB; 1HEZ; 18-JUL-03.
DR PDB; 1HKL; 12-MAR-97.
DR PDB; 1H7Z; 08-AUG-01.
DR PDB; 1M1M; 15-MAY-97.
DR Genew; HGNC:5716; IGKC.
DR MIM; 147200; -.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF000407; IGc1; 1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
FT NON_TER 1 1
FT DOMAIN 5 102 IG-LIKE.
FT DISULFID 26 86 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106 V -> L (IN INV(1,2) MARKER).
FT VARIANT 83 83 /FTID=VAR_003897.
FT CONFLICT 14 14 D -> N (IN REF. 5 AND 6).
FT CONFLICT 57 57 E -> Q (IN REF. 5 AND 6).
SQ SEQUENCE 106 AA; 11609 MW; 51984D1FDD372CE8 CRC64;

Query Match 4.3%; Score 148; DB 1; Length 106;
Best Local Similarity 29.5%; Pred. No. 0.0036;
Matches 28; Conservative 24; Mismatches 41; Indels 2; Gaps 1;

QY 525 POUYTLPPSRDELTLNGQVSLTCLVGFYPSDIAYEWESNG--OPENNYKTPPVLDSPGS 582
Db 5 PSVFIFPSDEQLSGTASVCLANFTYPRAKYQMKDANLQSGNSGSEVTBEDSKDST 64

QY 563 FFLYSKLTVDKSRMQGQNVFSCSYVMEALAHNYQ 617
Db 65 YLSSTLTLSKADYKHKVYACEVTHQGLSPVTK 99

RESULT 87
DCC_MOUSE
ID DCC_MOUSE STANDARD; PRT; 1447 AA.
AC P70211;
DT 01-NOV-1997 (Rel. 35, Created)
```



DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Tumor suppressor protein DCC precursor.  
 GN DCC.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).  
 RC STRAIN=BALB/c; TISSUE=Brain;  
 RX MEDLINE=96112625; PubMed=8570174;  
 RA Cooper H.M., Ames P., Brito J., Gad J., Wilks A.F.;  
 RT "Cloning of the mouse homologue of the deleted in colorectal cancer  
 gene (mdcc) and its expression in the developing mouse embryo.";  
 RL Oncogene 11:2243-2254(1995).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN=BALB/c; TISSUE=Brain;  
 RA Cooper H.M.;  
 RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=A;  
 CC IsoId=P70211-1; Sequence=Displayed;  
 CC Note=Isoform B is produced by alternative initiation at Met-85  
 CC of isoform A;  
 CC Name=C;  
 CC IsoId=P70211-2; Sequence=VSP\_002501;  
 CC Event=Alternative initiation;  
 CC Comment=2 isoforms, A (shown here) and B, are produced by  
 CC alternative initiation at Met-1 and Met-85;  
 CC -1- TISSUE SPECIFICITY: In the embryo, expressed at high levels in the  
 CC developing brain and neural tube. In adult, highly expressed in  
 CC brain with very low levels found in testis, heart and thymus.  
 CC -1- Isoform C is expressed only in the embryo.  
 CC -1- DEVELOPMENTAL STAGE: Low levels in early gestation. Highest levels  
 CC expressed during mid gestation. Levels decrease in late gestation  
 CC and remain at this level in the adult.  
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. DCC family.  
 CC -1- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 6 fibronectin type III domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [licenses@db-sib.ch](mailto:licenses@db-sib.ch)).  
 CC -----  
 CC EMBL: X85788; CAA59786.1; -.  
 DR HSSP: P56276; ITLK.  
 DR MGD: MGI:94869; DCC.  
 DR InterPro: IPR008957; FN\_III-like.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003962; FN\_III\_subd.  
 DR InterPro: IPR007110; Ig\_1-like.  
 DR InterPro: IPR003598; Ig\_C2.  
 DR Pfam: PF00041; Fn3; 6.  
 DR Pfam: PF00047; Ig; 4.  
 DR PRINTS: PR00014; FNTYPEIII.  
 DR SMART: SM00060; FN3; 6.  
 DR SMART: SM00408; IGC2; 3.  
 DR PROSITE: PS50835; IG\_LIKE; 4.  
 KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;  
 KW Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM A.  
 FT CHAIN 85 1447 TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM B.  
 FT INIT\_MET 85 85 FOR ISOFORM B.

FT DOMAIN 26 1097 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1098 1122 POTENTIAL.  
 FT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 133 135 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 139 229 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 234 326 IG-LIKE C2-TYPE 3.  
 FT DOMAIN 331 416 IG-LIKE C2-TYPE 4.  
 FT DOMAIN 426 522 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 525 618 FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 619 716 FIBRONECTIN TYPE-III 3.  
 FT DOMAIN 722 816 FIBRONECTIN TYPE-III 4.  
 FT DOMAIN 840 940 FIBRONECTIN TYPE-III 5.  
 FT DOMAIN 941 1042 FIBRONECTIN TYPE-III 6.  
 FT DISULFID 61 117 BY SIMILARITY.  
 FT DISULFID 161 212 BY SIMILARITY.  
 FT DISULFID 261 310 BY SIMILARITY.  
 FT DISULFID 352 400 BY SIMILARITY.  
 FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 819 838 Missing (in isoform C).  
 FT /FTId=VSP\_002501.  
 SQ SEQUENCE 1447 AA; 158298 MW; 0D1F1097C22D5B9F CRC64;  
 Query Match 4.3%; Score 148; DB 1; Length 1447;  
 Best Local Similarity 18.6%; Pred. No. 0.11;  
 Matches 121; Conservative 80; Mismatches 205; Indels 244; Gaps 28;  
 QY 34 GDVTELTCTASQKSIQFHKNKSNQIKILGNQSGFLTKGSKLND-RADSRSLMDQGNF 92  
 DB 154 GDVTLKCEVIGEGEMPTIHHQKNQD-----DNLPLRGDSRVVVLPSG-- 195  
 QY 93 PLIKIKIKIEDSPYICEVED-----QKEVQLD-----VFGLTANSDTHLLQSQ 137  
 DB 196 ALQISRLQPDSDGVRCSANPNASIRTNAGEAVERILSDPGLHRLQVFLQRPNVIAIEGK 255  
 QY 138 SLTL--TLESPPGS-----SPSVQCRSPRGKNIQSGKTLVSQLELOSSGTMTCVLYQ 188  
 DB 256 DAVLECCVSGVPPPSFTMLGEEVITQIRSK--VSLGGSSMLLSNVTDSDGYTTCVTVY 314  
 QY 189 NOKKVEFKIDIVLA---FOKASSIVYKKEGEQVEFSPPLAFVTEKLTGSGELMWQAEKRA 245  
 DB 315 KNEIISASALTYLVPPMFLNHSNLYAYESMIEF-----ECA 353  
 QY 246 SSSKSWITFD-LNKKEVSVKRTQDPRLQWKKLPLHLTLPLQALPOYAGSGNLTTL----- 299  
 DB 354 VSGKPVPTVMWKKGDV-----VIPSDFQIVGGSNLTILGVVK 392  
 QY 300 -----ALEATGKLGHOEVLVWRAIQLQKNLTCEVWGSPKMLSLKLENKAKV 351  
 DB 393 SDEGFYQCAVENAENAGNAQSSAQLIVPPALPSSSL-----PSAPRVLVLVSSRFVRL 447  
 QY 352 SKKEKPVVNLNPPAGMWQCLLSDGQVLESNIKVLPTWSTPVEPKSCDTHTCPCPCAP 411  
 DB 448 S-----WPPPAEAKNIQTF----- 463  
 QY 412 ELLGSGSVLFPKPKDITLMSITPEVTCVVDVSHEDPEVKENWYDGVVHNAKTKPR 471  
 DB 464 -----VF-FSRGDRERALTNTQGSLLQTLGVNLLKPEAMV----- 498  
 QY 472 EEOYNSITRYVSVLTVLHDMVNGKEYK-CRVSNK--ALPAPEKTISSAKQPREPOV 527  
 DB 499 -----TFRVVA-----YNEWGESQPIKAVATQPELQVGPVENL----- 534  
 QY 528 YTLPPSDELTKNOVSLTCLVGFPSDIAVENESQNPNNVKTTPPVLDSDGSPFLYS 587  
 DB 535 -----HAVSVS-----PISLITIE-----PPAY-ANGPVQGYR 562  
 QY 588 KLTIVDKSRMOQGN--VFSGSVMEAL--HNHYTQKSLSLSS--PGLQJDE 630



Db 563 LFCTEVTSGKEQNI EVDGSLGKLEGLKKFTETYLRLFLAVRYRGPSTDD 612

RESULT 88

SHS1\_HUMAN STANDARD; PRT; 503 AA.

AC P78334; 000683; Q43799; Q8N517; Q8TAL8; Q9H022; Q9UDX2; Q9UIJ6; Q9Y4U9;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DE 10-OCT-2003 (Rel. 42, Last annotation update)

DE Protein-tyrosine phosphatase non-receptor type substrate 1 precursor (SHP substrate-1) (SHPS-1) (inhibitory receptor SHPS-1) (Signal-regulatory protein alpha-1) (SIRP-alpha-1) (SIRP-alpha-2) (SIRP-alpha-3) (MYD-1 antigen) (Brain Ig-like molecule with tyrosine-based activation motifs) (Blt) (Macrophage fusion receptor) (p84).

GN PTPN1 OR SHPS1 OR SIRP OR MYD1 OR BIT OR MFR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OC NCBI\_TaxID=9606;

OX

RN (1)

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Brain;

RX MEDLINE=97223399; PubMed=9070220;

RA Yamao T., Matozaki T., Amano K., Matsuda Y., Takahashi N., Ochi F., Fujioka Y., Kasuga M.,

RT Mouse and human SHPS-1: molecular cloning of cDNAs and chromosomal localization of genes."

RL Biochem. Biophys. Res. Commun. 231:61-67(1997).

RN (2)

RP SEQUENCE FROM N.A. (ISOFORM 1), POLYMORPHISM, PHOSPHORYLATION, N-GLYCOSYLATION, AND INTERACTIONS WITH PTPN11; PTPN6 AND GRB2.

RC TISSUE=Placenta;

RX MEDLINE=97215901; PubMed=9062191;

RA Khartonenkov A., Chen Z., Sures I., Wang H., Schilling J., Ullrich A.;

RT "A family of proteins that inhibit signalling through tyrosine kinase receptors."

RL Nature 386:181-186(1997).

RN (3)

RP SEQUENCE FROM N.A. (ISOFORM 3), AND VARIANTS ASP-95; LEU-96; ASN-100; ARG-107; GLY-109 AND VAL-131.

RC TISSUE=Monocytes;

RX MEDLINE=98143722; PubMed=9485180;

RA Brooke G.P., Parsons K.R., Howard C.J.;

RT "Cloning of two members of the SIRP alpha family of protein tyrosine phosphatase binding proteins in cattle that are expressed on monocytes and a subpopulation of dendritic cells and which mediate binding to CD4 T cells."

RL Eur. J. Immunol. 28:1-11(1998).

RN (4)

RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS LEU-44; THR-50; THR-52; ARG-54; ALA-57; GLY-75; ASP-95; LEU-96; ASN-100; ARG-107; GLY-109 AND VAL-131.

RC TISSUE=Brain;

RX MEDLINE=20053880; PubMed=10585853;

RA Sano S.-I., Ohnishi H., Kubota M.;

RT "Gene structure of mouse BIT/SHPS-1."

RL Biochem. J. 344:667-675(1999).

RN (5)

RP SEQUENCE FROM N.A.

RX MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Gardner C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Grafham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Leharshaiho M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConachie L.J., McLeay K., Murray A.A., Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prichard S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Senta H.K., Showkeen R., Sims S., Skuse C.D., Smith M.L., Soderlund C., Stewart C.A., Sultson J.E., Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A., Wilmshurst L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.;

RT "The DNA sequence and comparative analysis of human chromosome 20."

RL Nature 414:865-871(2001).

RN (6)

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS LEU-44; THR-50; THR-52; ARG-54; ALA-57; GLY-75; ASP-95; LEU-96; ASN-100; ARG-107; GLY-109 AND VAL-131.

RC TISSUE=Brain, Kidney, and Skin;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klauener R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Scheele C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stadelman L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Useth T.B., Toshiyuki S., Cantucci P., Prange C., Rabe S.S., Lochellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.U., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN (7)

RP FUNCTION, AND INTERACTIONS WITH FYB; SCAP2 AND PTK2B.

RX MEDLINE=99401000; PubMed=10463599;

RA Timms J.F., Swanson K.D., Marie-Cardine A., Raab M., Rudd C.E., Schraven B., Neel B.G.;

RT "SHS-1 is a scaffold for assembling distinct adhesion-regulated multi-protein complexes in macrophages."

RL Curr. Biol. 9:927-930(1999).

RN (8)

RP PHOSPHORYLATION BY JAK2, AND INTERACTIONS WITH PTPN11 AND JAK2.

RX MEDLINE=20428742; PubMed=10842184;

RA Stofega M.R., Argersinger L.S., Wang H., Ullrich A., Carter-Su C.;

RT "Negative regulation of growth hormone receptor/JAK2 signaling by signal regulatory protein alpha."

RL J. Biol. Chem. 275:28222-28229(2000).

RN (9)

RP FUNCTION, AND INTERACTION WITH CD47.

RX MEDLINE=21400825; PubMed=11509594;

RA Latour S., Tanaka H., Demeure C., Mateo V., Rubio M., Brown E.J., Maliszewski C., Lindberg F.P., Oldenborg A., Ullrich A., Delapasse G., Sarfati M.;

RT "Bidirectional negative regulation of human T and dendritic cells by CD47 and its cognate receptor signal-regulator protein-alpha: down-regulation of IL-12 responsiveness and inhibition of dendritic cell activation."

RL J. Immunol. 167:2547-2554(2001).

CC -I- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts as docking protein and induces translocation of PTPN6, PTPN11 and other binding partners from the cytosol to the plasma membrane.







```

RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE-Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Feigold E.A., Grouse L.H., Derge J.G.,
RA Alteschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,
RA Stachenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallahy S.J.,
RA Rodriquez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Villalón D.K., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huliyk S.W.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywiński M.I., Skalska U., Smallus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP INTERACTION WITH TYROBP AND SYK.
RX PubMed=11169422;
RA Tomecello E., Cant C., Buehring H.-J., Vely F., Andre P., Seifert M.,
RA Ullrich A., Viver E.,
RT "Association of signal-regulatory proteins beta with KARAP/DAP-12."
RL Eur. J. Immunol. 30:2147-2156(2000).
RN [5]
RP INTERACTION WITH TYROBP, N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
RX MEDLINE=20072721; PubMed=10604985;
RA Dietrich J., Cella M., Seifert M., Buehring H.-J., Colonna M.,
RT "Signal-regulatory protein beta 1 is a DAP12-associated activating
RT receptor expressed in myeloid cells."
RL J. Immunol. 164:9-12(2000).
CC -1- FUNCTION: Immunoglobulin-like cell surface receptor involved in
CC the negative regulation of receptor tyrosine kinase-coupled
CC signaling processes. Participates also in the recruitment of
CC tyrosine kinase SYK.
CC -1- SUBUNIT: Interacts with TYROBP. This interaction results in the
CC recruitment of SYK.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named Isoforms=2;
CC Name=1;
CC IsoId=O00241-1; Sequence=displayed;
CC Name=2;
CC IsoId=O00241-2; Sequence=VSP_007026;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Detected in monocytes and dendritic cells.
CC -1- PTM: N-glycosylated.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y10376; CAA71404.1; -
CC EMBL; AL049634; CAB46661.2; -
CC EMBL; AL138804; CAC17540.1; -
CC EMBL; BC025286; AAH25286.1; -
CC GeneW; HGNC:115928; SIRPB1.
CC MIM; 603889; -
CC GO; GO:0005087; C:integral to plasma membrane; TAS.
CC GO; GO:0007166; P:cell surface receptor linked signal transdu. . .; TAS.

```

```

DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IgC1.2.
DR Repeat; PS50835; Ig_Like; 3.
KW Repeat; Signal; Transmembrane; Immunoglobulin domain; Glycoprotein;
KW Alternative splicing.
FT SIGNAL 1 26
FT CHAIN 27 398
FT DOMAIN 27 371
FT TRANSMEM 372 392
FT DOMAIN 393 398
FT DOMAIN 27 136
FT DOMAIN 147 246
FT DOMAIN 253 347
FT DISULFID 54 120
FT DISULFID 169 227
FT DISULFID 212 320
FT CARBOHYD 244 244
FT CARBOHYD 269 269
FT CARBOHYD 291 291
FT VARSPLIC 145 361
FT CONFLICT 53 53
FT CONFLICT 102 102
FT CONFLICT 229 229
FT CONFLICT 363 363
SQ SEQUENCE 398 AA; 43255 MW; A2AA08FEEB2BC52B CRC64;
Query Match 4.3%; Score 147; DB 1; Length 398;
Best Local Similarity 24.7%; Pred. No. 0.024; Mismatches 93; Indels 40; Gaps 10;
Matches 57; Conservative 41; Mismatches 93; Indels 40; Gaps 10;
QY 409 PABELGSPVFLPPEPKKDTLMSRTPEVTCVVVDSH--EDPEVKENWYGVGVANA 466
DB 147 PSAPVSGPAPV-----RATPEHTVSFTCEHSGSPRDIITKMFKNGLSDP 193
QY 467 KTK--PRE-----QNSTYRVSVTLVHOMLNKEKCKYSNLAALRA-PIEK--ISK 517
DB 194 QTNVDPADDSVYSIHSSTARVLTGRDVHSDVIT-----CEVAHTITLGGDPLRTANISE 247
QY 518 AKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTPPVL 577
DB 248 AIRVPTLELV-TQGMRAE---NQANVTGVSNFIPRGLQTLRLNGNSKRETSSTILE 303
QY 578 DSDGSFFLYSKLTVDKSRMQGQNVFSCVMEBALHNHYTOKSLSLSPQLQ 628
DB 304 NKDGTYNMMSWMLVNTCAHRDDVLTQVEHDG-----QDAVSKVALLEI 348
RESULT 90
SNE2 HUMAN STANDARD; PRT; 6885 AA.
AC Q8WXH0; Q8N1S3; Q8NF49; Q8TER7; Q8MWM3; Q8MWM4; Q8MWM5; Q8WXH1;
AC Q8N1S0; Q8N1F4; Q8Y2L4; Q8Y4R1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neuprin 2 (Nuclear envelope spectrin repeat protein 2) (Syn-2)
DE (Synaptic nuclear envelope protein 2) (Nucleus and actin connecting
DE element protein) (NUANCE protein).
GN SNE2 OR NDA OR KIA1011.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 9), FUNCTION, CHARACTERIZATION, AND
RP INTERACTION WITH F-ACTIN.
RX MEDLINE=22113122; PubMed=12118075;
RA Zhen Y.-Y., Libotte T., Munck M., Noegel A.A., Korenbaum E.,
RT "NUANCE, a giant protein connecting the nucleus and actin
RT cytoskeleton."

```



RL J. Cell Sci. 115:3207-3222(2002).  
 [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 4, 5 AND 7).  
 RX MEDLINE=21652858; PubMed=11792814;  
 RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Heyl L., Roberts R.G.,  
 Weissberg P.L., Ellis J.A., Shanahan C.M.;  
 RT "Neprins: a novel family of spectrin-repeat-containing proteins that  
 RT localize to the nuclear membrane in multiple tissues.";  
 RL J. Cell Sci. 114:4485-4498(2001).  
 [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=22296983; PubMed=12408964;  
 RA Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;  
 RT "The neprins are giant actin-binding proteins, orthologous to  
 RT Drosophila melanogaster muscle protein MSP-100.";  
 RL Genomics 80:473-481(2002).  
 [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 6).  
 RX TISSUE=Testis;  
 RA Pouskta A., Klein M., Mewes H.-W., Gaassenhuber J., Wiemann S.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=12508121;  
 RA Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,  
 Catellon L., Levy M., Barde V., De Berardinis V., Ureta-Vidal A.,  
 Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,  
 Sun H., Du H., Pepin K., Artiguenave F., Robert C., Cnaud C.,  
 Brule T., Jallouin O., Friedlander L., Samson G., Brocquier P.,  
 Cure S., Seguren B., Aniere F., Samain S., Crespeau H., Abbasi N.,  
 Alach N., Boescu D., Dickhoff R., Dots M., Dubois I., Friedman C.,  
 Gouvenoux M., James R., Madan A., Mairey-Bertrida B., Mangot S.,  
 Martin N., Menard M., Ozta S., Ratcliffe A., Shaffer T., Traak B.,  
 Vachier B., Bellemere C., Belser C., Besnard-Gonnet M.,  
 Batrol-Mavel D., Boulard M., Brier-Silla S., Combette S.,  
 Dufosse-Laurent V., Ferron C., Lechaplais C., Lousse C., Musset D.,  
 Magelant G., Patrau E., Petit E., Sirvain-Trukniewicz P., Trybou A.,  
 Vega-Cazay N., Bataille E., Bluet E., Borelatis I., Dubois M.,  
 Dumont C., Guerin T., Haffray S., Hamadi R., Munga J., Pellouin V.,  
 Robert D., Munderle B., Gauguet G., Roy A., Sainte-Marthe L.,  
 Verdier J., Verdier-Discala C., Hillier L.W., Fulton L., McPherson J.,  
 Maesuda F., Wilson R., Scarpelli C., Gyapay G., Winkler P., Saurin W.,  
 Queller F., Waterston R., Hood L., Weissbach J.;  
 RT "The DNA sequence and analysis of human chromosome 14.";  
 RL Nature 421:601-607(2003).  
 [6]  
 RP SEQUENCE FROM N.A. (ISOFORM 8).  
 RX TISSUE=Brain;  
 MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,  
 Staglestein M., Soares M.B., Bonaldo M.F., Cavaant T.L., Scheetz T.E.,  
 Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 Rana S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullah S.J.,  
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Halys S.W.,  
 Villalón D.K., Muzny K., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
 Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 Butlerfield V.S.N., Krzywinski M.T., Skalek U., Smallus D.E.,  
 Schmeier A., Schein J.E., Jones S.J.W., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [7]  
 RP SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 1-956 AND 5133-6885  
 RP FROM N.A.  
 RC TISSUE=Spleen, and Tongue;

RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.,  
 RA Nimomiya K., Magatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,  
 RA Katsura N., Saito K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishi S., Yamamoto Y., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suuki Y., Sugano S., Nagahari K., Masuo Y., Nagai K., Isogai T.;  
 RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 [8]  
 RP SEQUENCE OF 5709-6885 FROM N.A. (ISOFORM 2).  
 RP TISSUE=Brain;  
 RX MEDLINE=99246063; PubMed=10231032;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.,  
 RT "Prediction of the coding sequences of unidentified human genes. XIII.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 6:63-70(1999).  
 [9]  
 RP REVISIONS.  
 RC TISSUE=Brain;  
 RX MEDLINE=22158633; PubMed=12168954;  
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;  
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual  
 RT curation of 330 KIAA cDNA clones.";  
 RL DNA Res. 9:99-106(2002).  
 [10]  
 RP SEQUENCE OF 5754-6885 FROM N.A.  
 RX MEDLINE=2115917; PubMed=11230166;  
 RA Wiemann S., Weil B., Wellenreuther R., Gaassenhuber J., Glaesl S.,  
 RA Amoreux W., Boecker M., Bioecker H., Bauesachs S., Blum H.,  
 RA Lauber J., Dueserhoef A., Beyer A., Koehner K., Strack N.,  
 RA Mewes H.-W., Oetemaelder B., Obermaier B., Tampe J., Heubner D.,  
 RA Wambolt R., Korn B., Klein M., Pouskta A.;  
 RT "Towards a catalog of human genes and proteins: sequencing and  
 RT analysis of 500 novel complete protein coding human cDNAs.";  
 RL Genome Res. 11:422-435(2001).  
 [11]  
 RP FUNCTION: Involved in the maintenance of nuclear organization and  
 RP structural integrity. Probable anchoring protein which tethers the  
 RP nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton  
 RP by interacting with the nuclear envelope and with F-actin in the  
 RP cytoplasm.  
 [12]  
 RP SUBUNIT: Interacts with F-actin via its N-terminal domain.  
 [13]  
 RP SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The  
 RP largest part of the protein is cytoplasmic, while its C-terminal  
 RP part is associated with the nuclear envelope, most probably the  
 RP outer nuclear membrane. Remains associated with the nuclear  
 RP envelope during its breakdown in mitotic cells.  
 [14]  
 RP ALTERNATIVE PRODUCTS:  
 RP Event=Alternative splicing; Named isoforms=9;  
 RP Name=1;  
 RP IsoId=Q8WXH0-1; Sequence=Displayed;  
 RP Name=2;  
 RP IsoId=Q8WXH0-2; Sequence=VSP\_007164, VSP\_007166;  
 RP Note=No experimental confirmation available;  
 RP Name=3;  
 RP IsoId=Q8WXH0-3; Sequence=VSP\_007155;  
 RP Note=Produced by exon skipping that results in a frameshift. No  
 RP experimental confirmation available;  
 RP Name=4; Synonyms=Beta;  
 RP IsoId=Q8WXH0-4; Sequence=VSP\_007156;  
 RP Name=5; Synonyms=Alpha;  
 RP IsoId=Q8WXH0-5; Sequence=VSP\_007157, VSP\_007164, VSP\_007165;  
 RP Name=6;  
 RP IsoId=Q8WXH0-6; Sequence=VSP\_007158, VSP\_007165, VSP\_007166;  
 RP Note=No experimental confirmation available;  
 RP Name=7; Synonyms=Gamma;  
 RP IsoId=Q8WXH0-7; Sequence=VSP\_007154, VSP\_007163;  
 RP Name=8;  
 RP IsoId=Q8WXH0-8; Sequence=VSP\_007161, VSP\_007162;







DT 10-OCF-2003 (Ref. 42, last sequence update)  
DT 10-OCF-2003 (Ref. 42, last annotation update)  
DE Signal-regulatory protein beta-2 precursor (SIRP-beta-2) (SIRP-b2).  
GN SIRPb2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eultheria; Primates; Carnivora; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Placenta;  
RX MEDLINE=21036165; PubMed=1185750;  
RA Jones M., Scavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
RA Bleasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Coulson A., Coville G.J., Dearden R., Dhani P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,  
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.,  
RA Huckle E., Hunt A.R., Hunt S.E., Jekesch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
RA Lavelle A.M., Leverisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McCormack L.J., McLeay K., McMorris A.A.,  
RA Miness A.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
RA Phillimore B.J.C.T., Prichalingsam S.R., Plumb R.M., Ramsay H.,  
RA Rice C.M., Rose J.C.T., Scott C.E., Selha H.K., Showkenn R., Sims S.,  
RA Skuce C.D., Smith M.L., Soderlund C., Stewart C.A., Sultson J.E.,  
RA Swann R.M., Symeon N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
RA Whithead S.L., Whiteaker P., Willey D.L., Williams L., Williams S.A.,  
RA Wilmking L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
RA Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20.";  
RL Nature 414:865-871(2001).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Scapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Uddin T.B., Tohyiuki S., Caranci P., Prange C.,  
RA Raba S.S., Lonnellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buttefield J.S.N., Krzyzanski M.J., Skalska M., Smailus D.E.,  
RA Schmechel A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences";  
Proc. Natl. Acad. Sci. U.S.A. 99:15699-15903(2002).  
CC -1- FUNCTION: Probable immunoglobulin-like cell surface receptor.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

```
CC CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=1;  
CC IsoId=G9PIW8-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=G9PIW8-2; Sequence=VSP_007027;  
CC Note=No experimental confirmation available;  
CC Name=3;  
CC IsoId=G9PIW8-3; Sequence=VSP_007028;  
CC Note=No experimental confirmation available;  
CC -1- TISSUE SPECIFICITY: Detected in liver, and at very low levels in brain, heart, lung, pancreas, kidney, placenta and skeletal muscle.  
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements or send an email to license@ebi.ac.uk).  
-----  
CC EMBL; AB042624; BAA5692.1; -  
CC EMBL; ALI38804; CAC00474.1; -  
CC EMBL; BC020629; AAH20629.1; ALT_INIT.  
CC Genew; HGNC:15757; SIRPB2.  
CC MIM; 605466; -  
CC GO; GO:0007267; P:cell-cell signaling; TAS.  
CC GO; GO:0007242; P:intracellular signaling cascade; TAS.  
CC GO; GO:0008285; P:negative regulation of cell proliferation; TAS.  
CC InterPro; IPR007110; Ig-like.  
CC InterPro; IPR003597; Ig_c1.  
CC InterPro; IPR003006; Ig_MHC.  
CC InterPro; IPR003596; Ig_V.  
CC Pfam; PF00047; Ig; 3.  
CC SMART; SM00407; IGc1; 2.  
CC SMART; SM00406; IGV; 1.  
CC DR PROSITE; PS50835; IG_LIKE; 3.  
CC DR PROSITE; PS00290; IG_MHC; FALSE NEG.  
KW Repeat; Signal; Transmembrane; Immunoglobulin domain; Glycoprotein;  
KW Alternative splicing.  
FT SIGNAL 1 28 POTENTIAL.  
FT CHAIN 29 387 SIGNAL-REGULATORY PROTEIN BETA-2.  
FT DOMAIN 29 360 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 361 383 POTENTIAL.  
FT DOMAIN 384 387 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 29 137 IG-LIKE V-TYPE.  
FT DOMAIN 146 245 IG-LIKE C1-TYPE 1.  
FT DOMAIN 252 340 IG-LIKE C1-TYPE 2.  
FT DISULFID 53 119 POTENTIAL.  
FT DISULFID 168 226 POTENTIAL.  
FT DISULFID 271 329 POTENTIAL.  
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VASPLIC 1 33 Missing (in isoform 2).  
FT FTid=VSP_007027.  
FT VASPLIC 144 360 Missing (In isoform 3).  
FT FTid=VSP_007028.  
FT CONFLICT 263 263 A -> V (IN REF. 2).  
FT CONFLICT 286 286 L -> S (IN REF. 2).  
SQ SEQUENCE 387 AA; 42495 MW; F7F20C9F8B0E64B CRC64;  
  
Query Match 4.3%; Score 146; DB 1; Length 387;  
Best Local Similarity 24.8%; Pred. No. 0.027;  
Matches 57; Conservative 37; Mismatches 114; Indels 22; Gaps 8;
```



```
QY 465 NAKTKREQYNSTYRVSVLTVLHODWLNKGYCKENKALPA-PIEKT--ISKAKQ 521
DB 191 DQOTNDVDPGQSVANSIRSTARVLDPMVRSQVCEVAVHTLQGDPLEGTANLSAIRV 250
QY 522 PREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIIVAVESNGOPENNYKTPPVLDSPG 581
DB 251 PPTLEV-TQGPKR---AGQVAVTTCQVRKFFYQSLQTLTLENGVNCQRTASTLTLENKDG 306
QY 582 SFPLYSKLTVDKSRMOQGNVFCSCVMHEALHNHTQKSLSPGLQDFT 631
DB 307 TYNMTSWPLVNIISDQRDVVLTCQVGHDC-----QLAVSKRLALEVT 348

RESULT 93
LAC_CHICK
ID LAC_CHICK STANDARD; PRT; 103 AA.
AC P20763;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda chain C region.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87218480; PubMed=3107981;
RA Parvart R., Ziv E., Lerner F., Tel-Or S., Bursstein Y., Schechter I.;
RT "Analyses of chicken immunoglobulin light chain cDNA clones indicate
a few germline V lambda genes and allotypes of the C lambda locus.";
RC EMBO J. 6:97-102(1987).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04768; CA28461.1; -.
CC DR PIR; B26167; B26167.
CC DR HSSP; P01842; 7FAB.
CC DR InterPro; IPR007110; IG-1like.
CC DR InterPro; IPR003597; IG_c1.
CC DR InterPro; IPR003006; IG_MHC.
CC DR Pfam; PF00047; Ig_1.
CC DR SMART; SM00407; IGc1; 1.
CC DR PROSITE; PS50835; IG_LIKE; 1.
CC DR PROSITE; PS00290; IG_MHC; 1.
CC KW Immunoglobulin domain; Immunoglobulin C region.
CC FT NON TER 1 1
CC FT DOMAIN 6 99 IG-LIKE.
CC FT DISULFID 28 85
CC FT DISULFID 103 103 INTERCHAIN (WITH HEAVY CHAIN).
CC FT VARIANT 90 90 N -> D.
CC SEQUENCE 103 AA; 11361 MW; 77BF341B511B91B2 CRC64;

Query Match 4.3%; Score 145.5; DB 1; Length 103;
Best Local Similarity 32.2%; Pred. No. 0.005;
Matches 29; Conservative 19; Mismatches 39; Indels 3; Gaps 3;
```

```
RESULT 94
KACB_RAT
ID KACB_RAT STANDARD; PRT; 106 AA.
AC P01835;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig kappa chain C region, B allele.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Louvain;
RX MEDLINE=82082587; PubMed=6273908;
RA Sheppard H.W., Gutman G.A.;
RT "Allelic forms of rat kappa chain genes: evidence for strong
selection at the level of nucleotide sequence.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:7064-7068(1981).
RN [2]
RP SEQUENCE (BENCE-JONES PROTEIN S211).
RX MEDLINE=75212238; PubMed=807630;
RA Starace V., Querinjean P.;
RT "The primary structure of a rat kappa Bence Jones protein:
phylogenetic relationships of V- and C-region genes.";
RC J. Immunol. 115:59-62(1975).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC DR PIR; A93901; K1RFB.
CC DR HSSP; P01842; 2MCG.
CC DR InterPro; IPR007110; IG-1like.
CC DR InterPro; IPR003597; IG_c1.
CC DR InterPro; IPR003006; IG_MHC.
CC DR Pfam; PF00047; Ig_1.
CC DR SMART; SM00407; IGc1; 1.
CC DR PROSITE; PS50835; IG_LIKE; 1.
CC DR PROSITE; PS00290; IG_MHC; 1.
CC KW Immunoglobulin domain; Immunoglobulin C region.
CC FT NON TER 1 1
CC FT DOMAIN 5 102 IG-LIKE.
CC FT DISULFID 26 86
CC FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
CC FT CONFLICT 2 2 D -> N (IN REF. 2).
CC FT CONFLICT 30 30 N -> K (IN REF. 2).
CC FT CONFLICT 48 48 MISSING (IN REF. 2).
CC FT CONFLICT 79 79 E -> Q (IN REF. 2).
CC FT CONFLICT 87 87 E -> Q (IN REF. 2).
CC FT CONFLICT 98 98 V -> VM (IN REF. 2).
CC FT CONFLICT 100 100 S -> N (IN REF. 2).
CC SEQUENCE 106 AA; 11601 MW; 4CFA7CA820D1CA36 CRC64;

Query Match 4.2%; Score 145; DB 1; Length 106;
Best Local Similarity 31.5%; Pred. No. 0.0056;
Matches 29; Conservative 22; Mismatches 29; Indels 12; Gaps 2;
```

```
QY 525 POVYTLPPSRDELTKNOVSLTCLVKGFPSPDIIVAVESNGOPENNYKTPPVLDSPG 579
DB 5 PTVSIFPSTPQLATGASVAVCLMNNFYPRDISYKWKIDGTERRD-----GVLDVTFDQD 59
QY 580 --DGSFPLYSKLTVDKSRMOQGNVFCSCVMHEALHNHTQKSLSPGLQDFT 609
DB 60 SKDSTYSMSSTLSLKADYESHNLVTEVAK 91

RESULT 95
C166_CARAU
ID C166_CARAU STANDARD; PRT; 555 AA.
AC Q90304;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
```







RC TISSUE=Brain;  
 RX MEDLINE=92120663; PubMed=1769655;  
 RA Hlavayn M.L., Lemmon V.;  
 RT "Molecular structure and functional testing of human LICAM: an  
 RT interspecies comparison.";   
 RL Genomics 11:416-423(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92329299; PubMed=1627459;  
 RA Reid R.A., Hemperly J.J.;  
 RT "Variants of human L1 cell adhesion molecule arise through alternate  
 RT splicing of RNA.";   
 RL J. Mol. Neurosci. 3:127-135(1992).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Rosenthal A., Couelle O., Drescher B.;  
 RL Submitted (Apr-1994) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97432815; PubMed=9286655;  
 RA Brenner V., Nyakatura G., Rosenthal A., Platzner M.;  
 RT "Genomic organization of two novel genes on human Xq28: compact head  
 RT to head arrangement of IDH gamma and TRAP delta is conserved in rat  
 RT and mouse.";   
 RL Genomics 44:8-14(1997).  
 RN [6]  
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 RX MEDLINE=96147998; PubMed=9479034;  
 RA Couelle O., Nyakatura G., Tauden S., Elgar G., Brenner S.,  
 RA Platzner M., Drescher B., Jouet M., Kenwick S., Rosenthal A.;  
 RT "The neural cell adhesion molecule L1: genomic organisation and  
 RT differential splicing is conserved between man and the pufferfish  
 RT Fugu.";   
 RL Gene 208:7-15(1998).  
 RN [7]  
 RP SEQUENCE OF 20-36.  
 RX MEDLINE=88298876; PubMed=3136168;  
 RA Wolff J.M., Frank R., Mujoo K., Spiro R.C., Retsefeld R.A.,  
 RA Rathjen F.G.;  
 RT "A human brain glycoprotein related to the mouse cell adhesion  
 RT molecule L1.";   
 RL J. Biol. Chem. 263:11943-11947(1988).  
 RN [8]  
 RP SEQUENCE OF 332-371 FROM N.A.  
 RX MEDLINE=90353957; PubMed=2387585;  
 RA Djafarzadeh M., Mattei M.-G., Nguyen C., Roux D., Demergoot J.,  
 RA Denzot F., Moos M., Schachner M., Goridis C., Jordan B.R.;  
 RT "The gene encoding L1, a neural adhesion molecule of the  
 RT immunoglobulin family, is located on the X chromosome in mouse and  
 RT man.";   
 RL Genomics 7:587-593(1990).  
 RN [9]  
 RP SEQUENCE OF 353-1176 FROM N.A.  
 RX TISSUE=Fetal brain;  
 RX MEDLINE=92020233; PubMed=1923824;  
 RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;  
 RT "PCR walking from microdissection clone M54 identifies three exons  
 RT from the human gene for the neural cell adhesion molecule L1  
 RT (CAM-L1).";   
 RL Nucleic Acids Res. 19:5395-5401(1991).  
 RN [10]  
 RP SEQUENCE OF 809-1257 FROM N.A.  
 RX TISSUE=pancreas;  
 RX MEDLINE=92388257; PubMed=12477932;  
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemen G.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stabileton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.D., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahay J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";   
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [11]  
 RP SEQUENCE OF 1030-1257 FROM N.A.  
 RX MEDLINE=91132183; PubMed=1993895;  
 RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,  
 RA Stalcup W.B.;  
 RT "Isolation and sequence of partial cDNA clones of human L1: homology  
 RT of human and rodent L1 in the cytoplasmic region.";   
 RL J. Neurochem. 56:797-804(1991).  
 RN [12]  
 RP VARIANT HSAS TYR-264.  
 RX MEDLINE=94004956; PubMed=8401576;  
 RA Jouet M., Rosenthal A., Macfarlane J., Kenwick S., Donnai D.;  
 RT "A missense mutation confirms the L1 defect in X-linked hydrocephalus  
 RT (HSAS).";   
 RL Nat. Genet. 4:331-331(1993).  
 RN [13]  
 RP VARIANT HSAS/MASA LEU-1194.  
 RX MEDLINE=95187172; PubMed=7881431;  
 RA Franssen E., Schrander-Stumpel C., Vits L., Coucke P., van Camp G.,  
 RA Willems P.J.;  
 RT "X-linked hydrocephalus and MASA syndrome present in one family are  
 RT due to a single missense mutation in exon 28 of the LICAM gene.";   
 RL Hum. Mol. Genet. 3:2255-2256(1994).  
 RN [14]  
 RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.  
 RX MEDLINE=95004608; PubMed=7920659;  
 RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,  
 RA Paterson J., Metzenberg A., Ionaescu V., Temple K., Kenwick S.;  
 RT "X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked  
 RT hydrocephalus result from mutations in the L1 gene.";   
 RL Nat. Genet. 7:402-407(1994).  
 RN [15]  
 RP VARIANTS MASA GLN-210 AND ASN-598.  
 RX MEDLINE=95004609; PubMed=7920660;  
 RA Vits L., van Camp G., Coucke P., Franssen E., de Bouille K.,  
 RA Reyliens E., Korn B., Poustcka A., Wilson G., Schrander-Stumpel C.,  
 RA Winter R.M., Schwartz C., Willems P.J.;  
 RT "MASA syndrome is due to mutations in the neural cell adhesion gene  
 RT LICAM.";   
 RL Nat. Genet. 7:408-413(1994).  
 RN [16]  
 RP VARIANTS HSAS/MASA SER-9; SER-121; LYS-309; PHE-768; LEU-941 AND  
 RP CYS-1070.  
 RX MEDLINE=95282776; PubMed=7762552;  
 RA Jouet M., Moncla A., Paterson J., McKeown C., Fryer A., Carpenter N.,  
 RA Holmberg E., Madelin C., Kenwick S.;  
 RT "New domains of neural cell-adhesion molecule L1 implicated in  
 RT X-linked hydrocephalus and MASA syndrome.";   
 RL Am. J. Hum. Genet. 56:1304-1314(1995).  
 RN [17]  
 RP VARIANTS HSAS/MASA GLN-184, GLN-210, TYR-264; ARG-452; ASN-598 AND  
 RP LEU-1194.  
 RX MEDLINE=96153146; PubMed=8556302;  
 RA Franssen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;  
 RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,  
 RT retardation, adducted thumbs, spastic paraparesis and hydrocephalus  
 RT due to mutations in one single gene, L1.";   
 RL Eur. J. Hum. Genet. 3:273-284(1995).  
 RN [18]  
 RP ERRATUM.  
 RA Franssen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;  
 RL Eur. J. Hum. Genet. 4:126-126(1996).



```

RN [19]
RP VARIANTS HSAS/MASA/SPG1 SER-179 AND ARG-370.
RX MEDLINE=96057511; PubMed=7562969;
RA Ruiz J.C., Cuppens H., Legius E., Fryns J.-P., Glover T., Marynen P.,
CA Caesman J.-J.,
RT "Mutations in L1-CAM in two families with X linked complicated
RL epasitic paraplegia, MASA syndrome, and HSAS."
RL J. Med. Genet. 32:549-552(1995).
[20]
RN VARIANTS HSAS CVS-194 AND LEU-240.
RX MEDLINE=97083370; PubMed=8929944;
RA Gu S.-M., Orth U., Veske A., Enders H., Kluender K., Schloesser M.,
RA Engel W., Schwinger E., Gal A.;
RT "Five novel mutations in the L1CAM gene in families with X linked
RT hydrocephalus."
RL J. Med. Genet. 33:103-106(1996).
[21]
RN VARIANTS HSAS GLN-184; 439-VAL--THR-443 DEL; CVS-784 AND
RP 936-LEU--LEU-948 DEL.
RX MEDLINE=97338664; PubMed=9195224;
RA Macfarlane J.R., Du J.-S., Peyre M.E., Rameaden S., Donnai D.,
RA Charlton R., Garrett C., Tolmie J., Yates J.R.W., Berry C., Goudie D.,
RA Monica A., Hunt P., Hodgson S., Jouet M., Kenwick S.;
RT "Nine novel L1 CAM mutations in families with X-linked
RT hydrocephalus."
RL Hum. Mutat. 9:512-518(1997).
[22]
RN VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
RX MEDLINE=98180721; PubMed=9521424;
RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;
RT "Multiple exon screening using restriction endonuclease
RT fingerprinting (REF) detection of six novel mutations in the L1 cell
RT adhesion molecule (L1CAM) gene."

```

```

Query Match 4.2%; Score 144.5; DB 1; Length 1257;
Best Local Similarity 18.6%; Pred. No. 0.16;
Matches 129; Conservative 83; Mismatches 231; Indels 251; Gaps 32;

```

```

QY 8 RHLLLV-----LQALLPA--ATQGNKVLGKKDPVELTCTASQKSIQFMKSNQIKI 61
DB 409 RHGLLAAAYIVVQLPKILTDNQTYMVGSTAYILCAFGAPVSVOMLDEDTGV 468
QY 62 LGNGSFLTKGPSKLNDRADSRSLMDQGNFPLIINKIKEDSDTYICEVEDQKEEVQL 121
DB 469 LQDERFF-----PYANGTLGRDQANDTGRVFCFLANDQNNVTIM 509
QY 122 VFGLTANDTHLQ-----GSLTLTLES--PPSSSVQCRSRGNIQ----- 164
DB 510 A-NLKVKDQATQITQPRSTIEKKSGRVFTQASFPSPLOPSITWRGD-GRDLQELGDS 567
QY 165 -----GKTLVSQLELDQSGTWCTVLQNKQKVEFKIDIVLAF-----OKASSIYKK 214
DB 568 KYFIEDGR-LVTHSLDYSQGNYSCTVASTELDVESRQQLLVGSPGVPPLVUSDHL 626
QY 215 EGEQVESFPLA-----FTVEKLTSGLIMQAEARASSKSNITFDLNKKEVS 263
DB 627 TQSGRVAVMSPAEDHNAPIEKYDIEFEDKEMAPKMYSLGKPGVQSTITKLSPYHYT 686
QY 264 KRYVQDPKLGKGLPLHLTL--POLAPQ-----YASGNITLALBAKTKLHQBVLV 316
DB 687 FRVAVINKYGPGEPSVSETVVTPEAAPEKPPVDKVGNET-----TMMVI 733
QY 317 MRATQQLQNLTCVWGP-----TSPKMLSLKLENKAKVSKREKPVVPLNPEAGMMQ- 369
DB 734 -----TWKPLRMMDMNAPOVQYRVQ-----WRQGTGRGPQOE 765
QY 370 CLLSDSGVLLSNIKVLPTWSTPVEPKSCDHTCPCPAPELLGSPVFLFPKPKXT 429
DB 766 QIVSD--PFLVVSNTSTFVPEIKQAVNSQK-----GPE----- 799
QY 430 LMISTREVTGVVDVSHED-----PEYKFMVYDVGVVHN-----AKTKPR 471
DB 800 -----PQVT--IGYSGEDYPOALPE-----LEGIIINSSAVLVKMRPVDLAQVGH 844

```

```

QY 472 EEOYNSTRVSVLTLVHODPLNGKEYK-----CKVSNKALPA----- 509
DB 845 LRQNVNLY-----WREGSQKSKSHKHHDHVVVPANTTSVILSLRPSSYN 892
QY 510 -----PIEKTISKANGQPREPOVYTL-----PSRDELTKQVVS 543
DB 893 LEVOAFNGRSGPASEFTFSTPEGVGHPEALHLECOSNTSLLRMQPP-----LSHNGV- 947
QY 544 LTCLVKGFPSPDIAVENESNGQPENNYKTPPVTL 577
DB 948 LTGVLYSTHPLD-----EGCKQLSPNLRDPEL 975

RESULT 97
CAML_MOUSE
ID CAML_MOUSE STANDARD; PRT; 1260 AA.
AC P11627;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neural cell adhesion molecule L1 precursor (N-CAM L1).
GN L1CAM OR CAML1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RA MEDLINE=88318924; PubMed=3412448;
RA Moos M., Tacke R., Scherer H., Teplov D., Fruh K., Schachner M.;
RT "Neural adhesion molecule L1 as a member of the immunoglobulin
RT superfamily with binding domain similar to fibronectin."
RL Nature 334:701-703(1988).
CC -1- FUNCTION: Cell adhesion molecule with an important role in the
CC development of the nervous system. Involved in neuron-neuron
CC adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds
CC to axonin on neurons.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC L1/neurofascin/NG2CAM family.
CC -1- SIMILARITY: Contains 6 fibronectin-like C2-type domains.
CC -1- SIMILARITY: Contains 5 fibronectin type III domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, X12875; CAJ31368.1; -.
DR PIR, S05479; S05479.
DR HSP, P20241; 1CFB.
DR MGD, MGI:16721; L1cam.
DR GO, GO:0007411; P:axon guidance; IMP.
DR InterPro, IPR008957; FN_III-like.
DR InterPro, IPR003961; FN_III.
DR InterPro, IPR007110; IG-like.
DR InterPro, IPR003598; IG_C2.
DR Pfam, PF00041; fn3; 4.
DR Pfam, PF00047; ig; 6.
DR SMART, SM00060; FN3; 2.
DR SMART, SM00408; IGc2; 5.
DR PROSITE, PS50835; IG_LIKE; 6.
KW Neurogenesis; Cell adhesion; Developmental protein; Glycoprotein;
KW Transmembrane; Repeat; Immunoglobulin domain; Signal.
FT SIGNAL 1
FT CHAIN 20 1260 NEURAL CELL ADHESION MOLECULE L1.
FT DOMAIN 20 1123 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1124 1146 POTENTIAL.

```



```

FT DOMAIN 1147 1260 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 130 IG-LIKE C2-TYPE 1.
FT DOMAIN 138 225 IG-LIKE C2-TYPE 2.
FT DOMAIN 239 337 IG-LIKE C2-TYPE 3.
FT DOMAIN 332 419 IG-LIKE C2-TYPE 4.
FT DOMAIN 424 506 IG-LIKE C2-TYPE 5.
FT DOMAIN 517 600 IG-LIKE C2-TYPE 6.
FT DOMAIN 827 896 FIBRONECTIN TYPE-III 1.
FT DOMAIN 932 994 FIBRONECTIN TYPE-III 2.
FT DOMAIN 1032 1094 FIBRONECTIN TYPE-III 3.
FT SITE 553 555 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 57 113 BY SIMILARITY.
FT DISULFID 157 208 BY SIMILARITY.
FT DISULFID 263 311 BY SIMILARITY.
FT DISULFID 353 403 BY SIMILARITY.
FT DISULFID 447 496 BY SIMILARITY.
FT DISULFID 538 590 BY SIMILARITY.
FT CARBOHYD 100 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 875 875 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 968 968 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1260 AA; 140968 MM; 22B57001CB2A558 CR64;

Query Match 4.2%; Score 144.5; DB 1; Length 1260;
Best Local Similarity 19.9%; Pred. No. 0.16;
Matches 118; Conservative 89; Mismatches 218; Indels 167; Gaps 30;

QY 34 GPTVELTCTASQKSIQFHWKSNQIKILGNQSFITKGPSKLANDRARSRLMDQGNFP 93
DB 346 GPTARLDCCVQGRPOPEITWR-----INQSMETVKNQDKYRI-EGQS-- 387
QY 94 LIIKULKIEDSDTYICEVEDOK-----BEVOLLVGLTANSDTHL-LOGOSLTITL 143
DB 388 LILSNVQPTDVTWTOCEARNQGLLANAYIVVQPARILTKYDQTVAAVGGSTAYILLC 447
QY 144 ESPPGSPSPVQCRSPRGKNI-----QSGKTLVSQLELDQSGTWTCTVLAQNKYEF 195
DB 448 KAFGARVPVQWLDSEGTIVLODERFFPYANTLISRLQANDTGRYFOQANDQNTYI 507
QY 136 KIDIVVLAFOKASST-----VYKKEGEOVEFSPLAFTVEKLTSGSELMMQAEARASSK 249
DB 508 LANLQV---KEATQITQGRSAIEKKGARVFTQASFPSSL---QASITWRGDR----- 557
QY 250 SMITFDLKNKEYSVRKVNODPKLQGGKLLPLHLTPQALPOVAGSGNLTLMLEAKTKLH 309
DB 558 ----DLQGRGDSKYFIJEDGLTVI-----QSL-DYSDQANSVCVASTLDEVE 600
QY 310 QEVNVLVWKA-----TOLQKNLTCEWGP---TSPKLMSHLENKEAK 350
DB 601 SRAQLLVGSSPGVPHLELSDRHLKQSGVHLS---WSPADHNSPIEKYDIEFDKX-- 655
QY 351 VSKREKPVVNLPEKGMOCCLISDGOV---LLESNIKLLPTWSTPVEBKSCDKTHTCP 407
DB 656 -----MAPE--KMFSL-----GKVPNGNSTTLKLSPVVHYTFRVTAINKYGPSP 699

```

```

QY 408 CPAPELLGSPVFLFPPKDKDTLMISRPVEYTCVVVDVSHEDPEVK-----FNW 456
DB 700 SPVSSVVTTPA-----APEKN--PVDVRGSGENNNMNVITMKPLRMWDW 742
QY 457 YVDGEVHNNAKTPKEEOYNSTYRVSVLYTLHODMNGEKCKVSKALPAPIEKTIS 516
DB 743 NAPIQO-YRVQWRPQKQ--ETWRKQTV-----SDPLFVSNSTSFVPEIKVQ 788
QY 517 KAKQGR--EEOV---YT-----LPRSDLETKQVSLTGLVNGFYSDIA 557
DB 789 AVNNQGRKPEQVYITIGSGEDYPOVSPLEDTITFN-SSTVLVR-WRPVDLA 838

RESULT 98
KAC_MOUSE
ID_KAC_MOUSE STANDARD; PRT; 106 AA.
AC P01837;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig kappa chain C region.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE (MOPC 21).
RX MEDLINE=7305310; PubMed=4638343;
RA Svaeti J., Milstein C.;
RT "The complete amino acid sequence of a mouse kappa light chain.";
RL Biochem. J. 128:427-444(1972).
RN [2]
RP DISULFIDE BONDS (MOPC 21).
RX MEDLINE=73008889; PubMed=5073237;
RA Svaeti J., Milstein C.;
RT "The disulphide bridges of a mouse immunoglobulin G1 protein.";
RL Biochem. J. 126:837-850(1972).
RN [3]
RP SEQUENCE FROM N.A., AND REVISIONS TO 53-59.
RX MEDLINE=79084137; PubMed=103625;
RA Hamlyn P.H., Brownlee G.G., Cheng C.-C., Gait M.J., Milstein C.;
RT "Complete sequence of constant and 3' noncoding regions of an immunoglobulin mRNA using the dideoxynucleotide method of RNA sequencing.";
RL Cell 15:1067-1075(1978).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=82059477; PubMed=6170937;
RA Hamlyn P.H., Gait M.J., Milstein C.;
RT "Complete sequence of an immunoglobulin mRNA using specific priming and the dideoxynucleotide method of RNA sequencing.";
RL Nucleic Acids Res. 9:4485-4494(1981).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=81191915; PubMed=6262318;
RA Max E.E., Maizel J.V., Jr., Leder P.;
RT "The nucleotide sequence of a 5.5-kilobase DNA segment containing the mouse kappa immunoglobulin J and C region genes.";
RL J. Biol. Chem. 256:5116-5120(1981).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=81198949; PubMed=6785724;
RA Alendburger W., Neumaier P.S., Steinmetz M., Zachau H.G.;
RT "DNA sequence of the constant gene region of the mouse immunoglobulin kappa chain.";
RL Nucleic Acids Res. 9:971-981(1981).
RN [7]
RP SEQUENCE FROM N.A.
RX MEDLINE=88329081; PubMed=3138116;
RA de Waale P., Feys V., van de Voorde A., Molemans F., Piers W.;
RT "Expression in non-V-myeloid cells of mouse recombinant immunoglobulin directed against the tumour marker human placental alkaline phosphatase.";

```







```

FT      DEBITNHDGSPVNEBETPTPLTEPEKLPKEENGKEVLNA
FT      ETIEIKVNDIIOGKEDDIKA -> NCCGANGGNGGGMH
FT      LNAVGTPTVITMSISLCLF (in isoform short).
FT      /FTID=VSP_002390.
SQ      SEQUENCE      837 AA; 93203 MW; 70473B053A2D65A5 CRC64;
      Query Match      4.2%; Score 142; DB 1; Length 837;
      Best Local Similarity 17.7%; Pred. No. 0.13;
      Matches 101; Conservative 87; Mismatches 210; Indels 174; Gaps 22;

QY      3 RG-VPEPHLLLVLTQALLPRAATQGNKRVLAGKKGDTVELCTASQKKSIOGFHKNNSQIKI 61
DB      193 REIDPRDIIVIVNVPRAIMPOKSPNATARGEEMTLTCKASGSPDPTISFPRNG--KL 250
QY      62 LGNGSFLTKGPKSKLNDRAKSRSLMDQGNPFIITKNLKIESDPTICGV-----EDQKE 116
DB      251 ISENEKTYILKG-----SNTETLVNINIKKGGSVVCAKTKNKAQGDQK 293
QY      117 E-VOLLVFGLTANSPTHLQGS-----LTLTESPSPGSSPSVQC----- 155
DB      294 AFLQVAV-----QPHIIQLKNETTSENGHTVLVEAGEFVPEITWGRAIDGVWFSG 346
QY      156 -RSPRG---KNIOGKTLVSQLELDGSGTWCTVLQ---NOKKVEFKIDIVLAFOK 206
DB      347 DKSPDGRLEVKGQHGSSLHNRDVKLSDSGRYDCEAASRIQGHQRMHLDIEYAP-KFVS 405
QY      207 ASSIYVKEGGEQVBFPLAFVVEKLTSGGELMWQAEKSSSKSWITFDLKNKEVSVKRV 266
DB      406 NOTWYVSWEGNPINISCDVT-----ANPRAISIMHREKLL-----LPAKNT 446
QY      267 TDDPKLQMGKLPYLHLTPQALPQVAGSGNLTALAEAKTGKHQEVNLVVMBATQLOQKL 326
DB      447 TILKTHSVKRMILEIA-PTSDNDF---GRYCTATNRIGTFOEITELADV----- 495
QY      327 TCEVNGPPTSPKMLSLKLENKSAKYSKREKPVVNLPEAGMOCCLSDSGVLLSNIKY 386
DB      496 -----PSSPHGVKIIIEISQTKAKISFNK----- 518
QY      387 LPTWSPVPEKSGCDKHTHTCPRPAPRLLGPSVFLFPPEPK---DTLMISTPEVTCV 442
DB      519 -----PRSHGQVPIHHYQVDVKEVASETWIKVISHQVQTVV 554
QY      443 VDVSHDEPEVKFNWYDVGVVHNAKTKPREBQGNSTYRVVSVLTV-----LHQMVLNG 495
DB      555 V-LSSLEPPTY-----EIRVAAVNGGQGDYSKIEIFQTLVPRBPPRPISHGPPSG 606
QY      496 KEYKCVSNK-ALPAPIEKTISKAKGQPREPQ 526
DB      607 KSPKISITKQDDGCAPILEYIVKYRKSKEDEQ 638

RESULT 100
NCA2_HUMAN STANDARD; PRT; 761 AA.
AC P13592; P13593;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 1, 120 kDa isoform precursor (N-CAM 120)
DE (NCAM-120) (CD56 antigen).
GN NCAM1 OR NCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM N-CAM 120).
RC TISSUE=Skeletal muscle;
RX MEDLINE=89305258; PubMed=3253057;
RA Barton C.H., Dickson G., Gower H.J., Rowett L.H., Putt W.,
RA Elsom V., Moore S.E., Goridis C., Walsh F.S.;
RT "Complete sequence and in vitro expression of a tissue-specific
RT phosphatidylinositol-linked N-CAM isoform from skeletal muscle.",

```

```

RL      Development 104:165-173(1988).
RN [2]
RP SEQUENCE OF 491-761 FROM N.A. (ISOFORM N-CAM 120).
RC TISSUE=Skeletal muscle;
RX MEDLINE=87301755; PubMed=2887295;
RA Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
RT "Human muscle neural cell adhesion molecule (N-CAM): identification
RT of a muscle-specific sequence in the extracellular domain.";
RL Cell 50:1119-1130(1987).
RN [3]
RP SEQUENCE OF 491-655 FROM N.A. (ISOFORM C).
RX MEDLINE=89077552; PubMed=3203385;
RA Gower H.J., Barton C.H., Elsom V.L., Thompson J., Moore S.E.,
RA Dickson G., Walsh F.S.;
RT "Alternative splicing generates a secreted form of N-CAM in muscle
RT and brain.";
RL Cell 55:955-964(1988).
CC -! FUNCTION: This protein is a cell adhesion molecule involved in
CC neuron-neuron adhesion, neurite fasciculation, outgrowth of
CC neurites, etc.
CC -! SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=N-CAM 120;
CC IsoId=P13592-2; Sequence=Displayed;
CC Name=N-CAM 140;
CC IsoId=P13591-1; Sequence=External;
CC Name=C; Synonyms=Secreted;
CC IsoId=P13592-1; Sequence=VSP_002587;
CC -! SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -! SIMILARITY: Contains 2 fibronectin type III domains.
CC -! DATABASE: NAME=PROW; NOTE=CD guide CD56 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd56.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL: X16841; CAA34739.1; -
DR EMBL: M17409; AAS5912.1; -
DR EMBL: M22094; AAS5910.1; -
DR EMBL: M22092; AAS5911.1; -
DR EMBL: M22091; AAS5911.1; JOINED.
DR PIR: A31635; A31635.
DR PIR: S07784; ITHUNG.
DR Genew; HGNC:7656; NCAM1.
DR MIM; 116930; -
DR GO; GO:0016021; C:Integral to membrane; TAS.
DR GO; GO:0005865; C:Plasma membrane; TAS.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG_III-like.
DR InterPro; IPR003598; IG_C2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGc2; 5.
DR PROSITE; PS50835; IG-LIKE; 5.
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW GPI-anchor; Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 761 NEURAL CELL ADHESION MOLECULE 1, 120 kDa
FT ISOFORM.
FT DOMAIN 20 111 IG-LIKE C2-TYPE 1.
FT DOMAIN 116 205 IG-LIKE C2-TYPE 2.
FT DOMAIN 212 301 IG-LIKE C2-TYPE 3.
FT DOMAIN 308 403 IG-LIKE C2-TYPE 4.
FT DOMAIN 406 491 IG-LIKE C2-TYPE 5.

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:51:44 ; Search time 44.4955 Seconds  
(without alignments)  
4594.975 Million cell updates/sec

Title: SEQ3  
Perfect score: 3414  
Sequence: 1 NMRGVFRRHLVLQLALP.....DETCAADQGEJLDGWTTPD 648

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 125 summaries

Database :  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_ricket:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1714	50.2	397	6	009261 cercopithec
2	1710	50.1	397	6	009260 cercopithec
3	1709	50.1	397	6	095NE9 cercopithec
4	1705	49.9	397	6	009259 cercopithec
5	1705	49.9	397	6	009263 cercopithec
6	1698	49.7	397	6	009262 cercopithec
7	1697	49.7	397	6	002805 cercopithec
8	1610.5	47.2	457	6	08H2T8
9	1584.5	46.4	457	6	08H2T7
10	1277	37.4	470	4	0725W1
11	1266	37.1	469	4	0727P5
12	1265	37.1	509	4	08NF17
13	1264.5	37.0	482	4	072351
14	1258	36.8	479	4	096P08
15	1185.5	34.7	455	6	09XS78
16	1171	34.3	354	4	086TT2

17	1167	34.2	521	4	08N4Y9
18	1150	33.7	473	4	08TC63
19	1087	31.8	474	6	P79355
20	985	28.9	457	11	061396
21	957	28.0	433	11	055054
22	919	26.9	337	6	095M34
23	857	25.1	470	11	07TMR1
24	850.5	24.9	437	11	09R1A4
25	850.5	24.9	469	11	08R3V9
26	844	24.7	463	11	099LC4
27	840.5	24.6	473	11	099DL4
28	832	24.4	468	11	099LJ1
29	810.5	23.7	473	11	099LZ5
30	800.5	23.4	473	11	091Z05
31	799.5	23.4	474	11	08R3H6
32	417.5	12.2	613	4	096EY0
33	416	12.2	614	11	07TMR6
34	407.5	11.9	613	4	08WUK1
35	400	11.7	86	6	077596
36	400	11.7	86	6	077597
37	397	11.6	86	6	077594
38	397	11.6	86	6	077599
39	396.5	11.6	613	11	08VCX7
40	395	11.6	86	6	077595
41	392.5	11.5	597	4	096BB9
42	392	11.5	86	6	077598
43	389	11.4	614	4	096GAC
44	388	11.4	86	6	077601
45	383	11.2	86	6	077600
46	383	11.2	618	4	096AA6
47	378.5	11.1	597	4	09BUI0
48	376.5	11.0	597	4	09B088
49	373.5	10.9	588	4	08WUX4
50	357.5	10.5	375	4	086T11
51	357	10.5	71	4	013969
52	333.5	9.8	587	13	07TOR1
53	330.5	9.7	684	13	090544
54	305.5	8.9	99	6	029027
55	302.5	8.9	478	4	072379
56	298.5	8.7	494	4	096K68
57	296	8.7	496	4	096KX8
58	295	8.6	499	4	08NSK4
59	292	8.6	493	4	08NCL6
60	291	8.5	416	4	09NPP6
61	288	8.4	500	4	09BRV0
62	287.5	8.4	496	4	096DX0
63	286	8.4	492	4	072374
64	282	8.3	488	11	091WR1
65	281.5	8.2	384	4	09UP60
66	280.5	8.2	99	6	029028
67	279.5	8.2	497	4	08WY24
68	276.5	8.1	487	11	080217
69	274.5	8.0	487	11	099KX4
70	273	8.0	120	11	088650
71	272.5	8.0	489	11	08VCX4
72	271.5	8.0	482	11	090WB5
73	271.5	8.0	484	11	099LH6
74	270.5	7.9	486	11	091T07
75	270	7.9	481	11	091WT1
76	270	7.9	482	11	08K172
77	269.5	7.9	426	11	09DD09
78	266	7.8	481	11	091WT3
79	266	7.8	482	11	091W32
80	265.5	7.8	479	11	091WP5
81	264	7.7	484	11	08VEA0
82	264	7.7	488	11	08KOP2
83	261	7.6	480	11	091XE1
84	256.5	7.5	481	11	08VCV5
85	252	7.4	479	11	07TMR4
86	251.5	7.4	480	11	08K0Z4
87	250.5	7.3	479	11	099M22
88	240	7.0	487	13	09M6V7
89	237	6.9	90	11	P70443

08N4Y9 homo sapien  
08TC63 homo sapien  
P79355 felis silve  
061396 mus musculu  
055054 mus musculu  
095M34 equus cabal  
07TMR1 mus musculu  
09R1A4 mus musculu  
08R3V9 mus musculu  
099LC4 mus musculu  
099DL4 mus musculu  
099LJ1 mus musculu  
099LZ5 mus musculu  
091Z05 mus musculu  
08R3H6 mus musculu  
096EY0 homo sapien  
07TMR6 mus musculu  
08WUK1 homo sapien  
077596 mandrillu  
077597 mandrillu  
077594 cercopithec  
077599 theropithec  
08VCX7 mus musculu  
077595 cercocebus  
096BB9 homo sapien  
077598 papio sp. (  
096GAC homo sapien  
077601 lophocebus  
077600 lophocebus  
096AA6 homo sapien  
09BUI0 homo sapien  
09B088 homo sapien  
08WUX4 homo sapien  
086T11 homo sapien  
013969 homo sapien  
07TOR1 xenopus lae  
090544 gnglymasto  
029027 sus scrofa  
072379 homo sapien  
096K68 homo sapien  
096KX8 homo sapien  
08NSK4 homo sapien  
08NCL6 homo sapien  
09NPP6 homo sapien  
09BRV0 homo sapien  
096DX0 homo sapien  
072374 homo sapien  
091WR1 mus musculu  
09UP60 homo sapien  
029028 sus scrofa  
08WY24 homo sapien  
080217 mus musculu  
099KX4 mus musculu  
088650 marmota mon  
08VCX4 mus musculu  
090WB5 anas platyr  
099LH6 mus musculu  
091T07 mus musculu  
091WT1 mus musculu  
08K172 mus musculu  
09DD09 mus musculu  
091WT3 mus musculu  
091W32 mus musculu  
091WP5 mus musculu  
08VEA0 mus musculu  
08KOP2 mus musculu  
091XE1 mus musculu  
08VCV5 mus musculu  
07TMR4 mus musculu  
08K0Z4 mus musculu  
099M22 mus musculu  
09M6V7 gallus gall  
P70443 mus musculu



```

90      190      5.6      5636      4      Q96RW7      Q96rw7 homo sapien
91      186      5.4      268      13      Q90524      Q90524 ginglymosto
92      184.5      5.4      573      4      Q8WUJ8      Q8WUJ8 homo sapien
93      183.5      5.4      130      11      Q9D8W4      Q9D8W4 mus musculus
94      183      5.4      234      4      Q8N355      Q8N355 homo sapien
95      182.5      5.3      739      6      Q28260      Q28260 canis fam1
96      182.5      5.3      739      11      Q63669      Q63669 rattus norv
97      182      5.3      17352      5      Q95YM2      Q95YM2 procamburus
98      180.5      5.3      2693      5      Q8ISF3      Q8ISF3 caenorhabdi
99      180.5      5.3      2708      5      Q8ISF4      Q8ISF4 caenorhabdi
100      180.5      5.3      2780      5      Q8MNS0      Q8MNS0 caenorhabdi
101      180.5      5.3      2808      5      Q8MNS1      Q8MNS1 caenorhabdi
102      180.5      5.3      18519      5      Q8ISF6      Q8ISF6 caenorhabdi
103      180.5      5.3      18534      5      Q8ISF7      Q8ISF7 caenorhabdi
104      180      5.3      235      11      Q99M11      Q99M11 mus musculus
105      179      5.2      1031      13      Q90VM2      Q90VM2 brachydanio
106      178.5      5.2      2673      4      Q96SC3      Q96SC3 homo sapien
107      178.5      5.2      6620      4      Q96AA2      Q96AA2 homo sapien
108      177.5      5.2      4162      13      Q98918      Q98918 gallus gall
109      176.5      5.2      233      4      Q96169      Q96169 homo sapien
110      176.5      5.2      259      13      Q90530      Q90530 ginglymosto
111      176.5      5.2      1323      13      Q08476      Q08476 gallus gall
112      175.5      5.1      2053      4      Q8WUJ8      Q8WUJ8 homo sapien
113      175.5      5.1      2113      4      Q8TD84      Q8TD84 homo sapien
114      174.5      5.1      8081      5      Q7120      Q7120 caenorhabdi
115      174      5.1      236      4      Q96B61      Q96B61 homo sapien
116      174      5.1      650      6      Q9GKR2      Q9GKR2 bos taurus
117      174      5.1      739      6      Q9GKR3      Q9GKR3 bos taurus
118      173      5.1      34350      4      Q8WZ42      Q8WZ42 homo sapien
119      172.5      5.1      234      4      Q72ZU7      Q72ZU7 homo sapien
120      172.5      5.1      2053      4      Q81ZY4      Q81ZY4 homo sapien
121      172.5      5.1      7962      4      Q10465      Q10465 homo sapien
122      171.5      5.0      240      4      Q8WUK3      Q8WUK3 homo sapien
123      171.5      5.0      257      13      Q90536      Q90536 ginglymosto
124      171      5.0      233      4      Q8N5F4      Q8N5F4 homo sapien
125      170      5.0      237      13      Q90545      Q90545 ginglymosto

```

## ALIGNMENTS

```

RESULT 1
009261      PRELIMINARY;      PRT;      397 AA.
AC      009261;
DT      01-JUL-1997 (TReMBLrel. 04, Created)
DT      01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT      01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE      CD4 (Fragment).
GN      CD4.
OS      Cercopithecus sabaenus.
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC      Cercopithecinae; Cercopithecus.
OX      NCBI_TaxID=60711;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98017879; PubMed=9379478;
RA      Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA      Corbet S., Barre-Sinoussi F., Allan J.S.;
RA      "Relation between phylogeny of African green monkey CD4 genes and
RA      their respective simian immunodeficiency virus genes.";
RL      J. Med. Primatol. 26:120-128(1997).
DR      EMBL; AF001225; AAB60872.1; -
DR      HSSP; P01730; IWIQ.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0006955; P:immune response; IEA.
DR      InterPro; IPR000973; CD4_TcAg.
DR      InterPro; IPR007110; IG-like.
DR      InterPro; IPR003596; IG_v.
DR      Pfam; PF00047; Ig_2.
DR      PRINTS; PR00692; CD4TCANTIGEN.
DR      SMART; SM00406; IGV; 1.

```

```

DR      PROSITE; PSS0835; IG_LIKE; 1.
FT      NON_TER      1
FT      NON_TER      397
SQ      SEQUENCE      397 AA; 43945 MW; 80C963B92A68CD3 CRC64;
Query Match      50.2%; Score 1714; DB 6; Length 397;
Best Local Similarity 89.7%; Pred. No. 3,6e-122;
Matches 331; Conservative 16; Mismatches 22; Indels 0; Gaps 0;
QY      28      VLIGKKDTELCTSTQKSIQPHMKNQIKILGQSGFLTKGPKLNDRAISRSLW      87
Db      1      VLIGKKDTELCTSTQKSIQPHMKNQIKILGQSGFLTKGSKLNDRIDSRKSLW      60
QY      88      DQGNFPLINKLIEDSDTYICEVEDQKEVOLVFLGLTNSPDTHLQGSGLTTLTSSPP      147
Db      61      DQGFSSMIINKLIEDSEYICEVENKEVELLVFLGLTNSPDTHLQGSGLTTLTSSPP      120
QY      148      GSSPSVOCRRPKNIGGKTLVSQLELDSGTTVTVLQNKVPEFKIDIVVLAQKA      207
Db      121      GSSPSVCRSPRKNIGGKTLVSQLELDSGTTVTVLQNKVPEFKIDIVVLAQKA      180
QY      208      SSIVYKKEGROVERSPFLATFVEKLTGSGELMMQAEPASSKSWITFDLNKEVSVKRYT      267
Db      181      SSIVYKKEGROVERSPFLATFVEKLTGSGELMMQAEPASSKSWITFDLNKEVSVKRYT      240
QY      268      QDPRLQWKKLPLHLTLPOALPOYAGSGNLTALAEATGKLHDEVNLVWRATQLQKRLT      327
Db      241      QDPRLQWKKLPLHLTLPOALPOYAGSGNLTALAEATGKLHDEVNLVWRATQLQKRLT      300
QY      328      CEWGPSPKLMISLKENKAESKREKPVWVLANPAGMWOCILSSGOVLLESNIKVL      387
Db      301      CEWGPSPKLMISLKENKAESKREKPVWVLANPAGMWOCILSSGOVLLESNIKVL      360
QY      388      PTWSTPVEP      396
Db      361      PTWSTPVEP      369

```

```

RESULT 2
009260      PRELIMINARY;      PRT;      397 AA.
AC      009260;
DT      01-JUL-1997 (TReMBLrel. 04, Created)
DT      01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT      01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE      CD4 (Fragment).
GN      CD4.
OS      Cercopithecus sabaenus.
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC      Cercopithecinae; Cercopithecus.
OX      NCBI_TaxID=60711;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98017879; PubMed=9379478;
RA      Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA      Corbet S., Barre-Sinoussi F., Allan J.S.;
RA      "Relation between phylogeny of African green monkey CD4 genes and
RA      their respective simian immunodeficiency virus genes.";
RL      J. Med. Primatol. 26:120-128(1997).
DR      EMBL; AF001224; AAB60871.1; -
DR      HSSP; P01730; IWIQ.
DR      GO; GO:0006955; P:immune response; IEA.
DR      GO; GO:0016020; C:membrane; IEA.
DR      InterPro; IPR000973; CD4_TcAg.
DR      InterPro; IPR007110; IG-like.
DR      InterPro; IPR003596; IG_v.
DR      Pfam; PF00047; Ig_2.
DR      PRINTS; PR00692; CD4TCANTIGEN.
DR      SMART; SM00406; IGV; 1.
DR      PROSITE; PSS0835; IG_LIKE; 1.
FT      NON_TER      1
FT      NON_TER      397

```



SQ SEQUENCE 397 AA; 43882 MW; 478BB277E992EE89 CRC64;  
 Query Match 50.1%; Score 1710; DB 6; Length 397;  
 Best Local Similarity 89.4%; Pred. No. 7, 3e-122;  
 Matches 330; Conservative 17; Mismatches 22; Indels 0; Gaps 0;

QY 28 VVLGKKGDTVELTCTASQKSIQPHWKNNOIKILGNQGSFLTKGSPSKLNDRAISRSLW 87  
 DB 1 VVLGKKGDTVELTCTASQKSIQPHWKNNOIKILGNQGSFLTKGSPSKLNDRAISRSLW 60  
 QY 88 DQGNPFLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDDLLOGOSITLTLESPP 147  
 DB 61 DQGCSSMIITKNLKIEDSDTYICEVENKEVEBELVFGLTANSDDLLOGOSITLTLESPP 120  
 QY 148 GSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIIVLAFQKA 207  
 DB 121 GSSPSVKCRSPRGKNIQGGRTLSVPLEKODSGTWTCTVSDQNTVEFKIDIIVLAFQKA 180  
 QY 208 SSIYVKKEGEQVEFSPPLAFVTEKLTGSGELMWQERASSSSKSWITFDLKNKEVSKVKT 267  
 DB 181 SSTYVKKEGEQVEFSPPLAFVTEKLTGSGELMWQERASSSSKSWITFDLKNKEVSKVKT 240  
 QY 268 QDPKLOMGKKLPLHLTLPLQALPOYAGSGNLTALAEKTKGLHQBVLVVMRATQLOKNI 327  
 DB 241 QDPKLOMGKKLPLHLTLPLQALPOYAGSGNLTALAEKTKGLHQBVLVVMRATQLOKNI 300  
 QY 328 CEVWGPTSPKMLSLKLENKEAKVSKREKPVVWVLPBAGMOCCLISDQGVLLSENIKYL 387  
 DB 301 CEVWGPTSPKMLSLKLENKEAKVSKREKPVVWVLPBAGMOCCLISDQGVLLSENIKYL 360  
 QY 388 PTWSTPVEP 396  
 DB 361 PTWSTPVEP 369

RESULT 3  
 Q95NE9 PRELIMINARY; PRT; 397 AA.  
 ID Q95NE9;  
 AC Q95NE9;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE CD4 (Fragment).  
 GN CD4.  
 OS Cercopithecus pygerythrus.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;  
 OC Cercopithecinae; Cercopithecus.  
 OX NCBI\_TaxID=60710;  
 RX MEDLINE=98017679; PubMed=9379478;  
 RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,  
 Corbet S., Barre-Sinoussi F., Allan J.S.;  
 RT "Relation between phylogeny of African green monkey CD4 genes and  
 their respective simian immunodeficiency virus genes.";  
 RL J. Med. Primatol. 26:120-128(1997).  
 DR EMBL; AF001227; AAB60874.1; -;  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0006955; P:immune response; IEA.  
 DR InterPro; IPR000973; CD4\_TcAg.  
 DR InterPro; IPR007110; IG\_Like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; Ig\_2.  
 DR SMART; PR00692; CD4TCANTIGEN.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PSS0835; IG\_LIKE; 1.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 397 AA; 43946 MW; 21C3E30882ABFBC0 CRC64;

Query Match 50.1%; Score 1709; DB 6; Length 397;  
 Best Local Similarity 89.4%; Pred. No. 8, 7e-122;

Matches 330; Conservative 16; Mismatches 23; Indels 0; Gaps 0;

QY 28 VVLGKKGDTVELTCTASQKSIQPHWKNNOIKILGNQGSFLTKGSPSKLNDRAISRSLW 87  
 DB 1 VVLGKKGDTVELTCTASQKSIQPHWKNNOIKILGNQGSFLTKGSPSKLNDRAISRSLW 60  
 QY 88 DQGNPFLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDDLLOGOSITLTLESPP 147  
 DB 61 DQGCSSMIITKNLKIEDSDTYICEVENKEVEBELVFGLTANSDDLLOGOSITLTLESPP 120  
 QY 148 GSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIIVLAFQKA 207  
 DB 121 GSSPSVKCRSPRGKNIQGGRTLSVPLEKODSGTWTCTVSDQNTVEFKIDIIVLAFQKA 180  
 QY 208 SSIYVKKEGEQVEFSPPLAFVTEKLTGSGELMWQERASSSSKSWITFDLKNKEVSKVKT 267  
 DB 181 SSTYVKKEGEQVEFSPPLAFVTEKLTGSGELMWQERASSSSKSWITFDLKNKEVSKVKT 240  
 QY 268 QDPKLOMGKKLPLHLTLPLQALPOYAGSGNLTALAEKTKGLHQBVLVVMRATQLOKNI 327  
 DB 241 QDPKLOMGKKLPLHLTLPLQALPOYAGSGNLTALAEKTKGLHQBVLVVMRATQLOKNI 300  
 QY 328 CEVWGPTSPKMLSLKLENKEAKVSKREKPVVWVLPBAGMOCCLISDQGVLLSENIKYL 387  
 DB 301 CEVWGPTSPKMLSLKLENKEAKVSKREKPVVWVLPBAGMOCCLISDQGVLLSENIKYL 360  
 QY 388 PTWSTPVEP 396  
 DB 361 PTWSTPVEP 369

RESULT 4  
 Q09259 PRELIMINARY; PRT; 397 AA.  
 ID Q09259;  
 AC Q09259;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE CD4 (Fragment).  
 GN CD4.  
 OS Cercopithecus sabaues.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;  
 OC Cercopithecinae; Cercopithecus.  
 OX NCBI\_TaxID=60711;  
 RX MEDLINE=98017879; PubMed=9379478;  
 RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,  
 Corbet S., Barre-Sinoussi F., Allan J.S.;  
 RT "Relation between phylogeny of African green monkey CD4 genes and  
 their respective simian immunodeficiency virus genes.";  
 RL J. Med. Primatol. 26:120-128(1997).  
 DR EMBL; AF001223; AAB60870.1; -;  
 DR HSP; P01730; IWO.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0006955; P:immune response; IEA.  
 DR InterPro; IPR00973; CD4\_TcAg.  
 DR InterPro; IPR007110; IG\_Like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; Ig\_2.  
 DR SMART; PR00692; CD4TCANTIGEN.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PSS0835; IG\_LIKE; 1.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 397 AA; 43881 MW; 7CE39AD0F8506C81 CRC64;

Query Match 49.9%; Score 1705; DB 6; Length 397;  
 Best Local Similarity 89.2%; Pred. No. 1, 8e-121;  
 Matches 329; Conservative 18; Mismatches 22; Indels 0; Gaps 0;

QY 28 VVLGKKGDTVELTCTASQKSIQPHWKNNOIKILGNQGSFLTKGSPSKLNDRAISRSLW 87



```
Db 1 VVLGKGGDTVELTCNASQKTTQFFHMKNSNOIKILGQGSFLTKGSKLDRIDSRKSLW 60
Qy 88 DQGNFPLIKNLKIEDSDTYICEVEDOKEEVOLLVFGLTANSPTHLLOGSLTLTLESPP 147
Db 61 DQGFCSWIIKNLKIETSETYICEVENKEEVEVLVFGLTANSPTHLLOGSLTLTLESPP 120
Qy 148 GSSPSVQCRSPRGKNIQGGKTLISVQLBLQDSGTWCTVLOKQKVEFKIDIVVLAPOKA 207
Db 121 GSSPSVQCRSPRGKNIQGGKTLISVQLBLQDSGTWCTVLOKQKVEFKIDIVVLAPOKA 180
Qy 208 SSIYKKEGQVEFSFPFLATFVEKLTGSGELMWAERASSSKSWITPDLKNKEVSVKRYT 267
Db 181 SSIYKKEGQVEFSFPFLATFVEKLTGSGELMWAERASSSKSWITPDLKNKEVSVKRYT 240
Qy 268 QDPKLQWKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHOEVNLVWMRATQFQENLT 327
Db 241 QDPKLQWKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHOEVNLVWMRATQFQENLT 300
Qy 328 CEVWGPTSPKLMSTKLENKEAKYSKREKPYVNLNPEAGWQCLSDSGVLLSNIKVL 387
Db 301 CEVWGPTSPKLMSTKLENKEAKYSKREKPYVNLNPEAGWQCLSDSGVLLSNIKVL 360
Qy 388 PTWSTPVEP 396
Db 361 PTWPTPVQP 369
```

```
RESULT 5
ID 009263 PRELIMINARY; PRT; 397 AA.
AC 009263;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE CD4 (Fragment) .
GN CD4.
OS Cercopithecus tantalus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=60712;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001222; AAB60869.1; -.
DR HSSP; P01730; IWIQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR00973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43994 MW; A3CD031535A51524 CRC64;
```

Query Match 49.9%; Score 1705; DB 6; Length 397;  
Best Local Similarity 89.2%; Pred. No. 1.8e-121;  
Matches 329; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

```
Qy 28 VVLGKGGDTVELTCNASQKSIQFHWKNSNOIKILGQGSFLTKGSKLDRIDSRKSLW 87
Db 1 VVLGKGGDTVELTCNASQKTTQFFHMKNSNOIKILGQGSFLTKGSKLDRIDSRKSLW 60
```

```
Qy 88 DQGNFPLIKNLKIEDSDTYICEVEDOKEEVOLLVFGLTANSPTHLLOGSLTLTLESPP 147
Db 61 DQGFCSWIIKNLKIETSETYICEVENKEEVEVLVFGLTANSPTHLLOGSLTLTLESPP 120
Qy 148 GSSPSVQCRSPRGKNIQGGKTLISVQLBLQDSGTWCTVLOKQKVEFKIDIVVLAPOKA 207
Db 121 GSSPSVQCRSPRGKNIQGGKTLISVQLBLQDSGTWCTVLOKQKVEFKIDIVVLAPOKA 180
Qy 208 SSIYKKEGQVEFSFPFLATFVEKLTGSGELMWAERASSSKSWITPDLKNKEVSVKRYT 267
Db 181 SSIYKKEGQVEFSFPFLATFVEKLTGSGELMWAERASSSKSWITPDLKNKEVSVKRYT 240
Qy 268 QDPKLQWKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHOEVNLVWMRATQFQENLT 327
Db 241 QDPKLQWKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHOEVNLVWMRATQFQENLT 300
Qy 328 CEVWGPTSPKLMSTKLENKEAKYSKREKPYVNLNPEAGWQCLSDSGVLLSNIKVL 387
Db 301 CEVWGPTSPKLMSTKLENKEAKYSKREKPYVNLNPEAGWQCLSDSGVLLSNIKVL 360
Qy 388 PTWSTPVEP 396
Db 361 PTWPTPVQP 369
```

```
RESULT 6
ID 009262 PRELIMINARY; PRT; 397 AA.
AC 009262;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE CD4 (Fragment) .
GN CD4.
OS Cercopithecus tantalus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=60712;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001221; AAB60868.1; -.
DR HSSP; P01730; IWIQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR00973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43954 MW; CF7F2F5D82335B0D CRC64;
```

Query Match 49.7%; Score 1698; DB 6; Length 397;  
Best Local Similarity 89.2%; Pred. No. 6e-121;  
Matches 329; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

```
Qy 28 VVLGKGGDTVELTCNASQKSIQFHWKNSNOIKILGQGSFLTKGSKLDRIDSRKSLW 87
Db 1 VVLGKGGDTVELTCNASQKTTQFFHMKNSNOIKILGQGSFLTKGSKLDRIDSRKSLW 60
Qy 88 DQGNFPLIKNLKIEDSDTYICEVEDOKEEVOLLVFGLTANSPTHLLOGSLTLTLESPP 147
Db 61 DQGFCSWIIKNLKIETSETYICEVENKEEVEVLVFGLTANSPTHLLOGSLTLTLESPP 120
```



```

QY 148 GSSPSVQCRSPRGKNIQGKTLVSQLELDSDGTCTCTVLQNOCKVFEKIDIVLAFOKA 207
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GSSPSVKCRSPRGKNIQGRRTLSVPLEERQDSGTCTCTVSDQNTVEFKIDIVLAFOKA 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 208 SSIYVKEGEQVEFSPFLAFVTEKLTGSGELMWQAEKSSKSWITFDLKNKEVSVKQVT 267
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SSTYVKEGEQVEFSPFLAFVTEKLTGSGELMWQAEKSSKSWITFDLKNKEVSVKQVT 240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 268 QDPKLOMGKRLPLHLTLPOALPOYAGSGNLTALAEKTKLHGEVNLVVMRATQLOKNT 327
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 QDPKLOMGKRLPLHLTLPOALPOYAGSGNLTALAEKTKLHGEVNLVVMRATQLOKNT 300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 328 CEVWGPTSPKMLSLKLENKKAIVSKREKPVWVNLNPEAGMOCCLSDSGQVLLSNIKVL 387
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 CEVWGPTSPKMLSLKLENKKAIVSKQAKAVWVNLNPEAGMOCCLSDSGQVLLSNIKVL 360
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 388 PTWSTPVP 396
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 PTWSTPVP 369
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 7

```

ID 002805 PRELIMINARY; PRT; 397 AA.
AC 002805: 07593;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3)
    (Fragment).
GN CD4.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OC NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood;
RA MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Mueller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RA "Rejection between phylogenetic and African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
RN [2]
RP SEQUENCE OF 80-165 FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH p56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 3 C2-LIKE DOMAINS.
DR EMBL: AF001226; AAB60873.1; -
DR EMBL: AF057380; AAC25124.1; -
DR HSBP: P01730; IWI0.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0006955; P:immune response; IEA.
DR InterPro: IPR000973; CD4 TCAG.
DR InterPro: IPR007110; IG-Like.
DR InterPro: IPR003596; IG_v.
DR Pfam: PF00047; I9; 2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV; 1.
DR Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Lipoprotein; Palmitate; Repeat.

```

```

FT NON TER 1 1
FT DOMAIN <1 370
FT TRAMEM 371 391
FT DOMAIN 392 >397
FT DOMAIN <1 98
FT DOMAIN 99 176
FT DOMAIN 177 290
FT DOMAIN 291 347
FT CAROHD 15 15
FT CAROHD 30 30
FT CAROHD 269 269
FT CAROHD 298 298
FT DISULFD 14 82
FT DISULFD 128 157
FT DISULFD 301 343
FT LIPID 392 392
FT LIPID 395 395
FT NON TER 397 397
SQ SEQUENCE 397 AA; 43980 MW; F74C42E22B196155 CRC64;

```

Query Match 49.7%; Score 1697; DB 6; Length 397;  
 Best Local Similarity 88.9%; Pred. No. 7.2e-121;  
 Matches 328; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

```

QY 28 VVLGKGGDTVELTCTASQKSIQPHMKNSNOIKILGNQSFETGKPSKLNDRADSRSLM 87
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VVLGKGGDTVELTCTASQKTIQPHMKNSNOIKILGNQSFETGKPSKLNDRADSRSLM 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 88 DQGNPFLIKKLIKEDSDTYICEVEDQEEVQLVFGLTANSDFHLLQGOSLTLLTESPP 147
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 DQGNPFLIKKLIKEDSDTYICEVEDQEEVQLVFGLTANSDFHLLQGOSLTLLTESPP 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 148 GSSPSVQCRSPRGKNIQGKTLVSQLELDSDGTCTCTVLQNOCKVFEKIDIVLAFOKA 207
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GSSPSVKCRSPRGKNIQGRRTLSVPLEERQDSGTCTCTVSDQNTVEFKIDIVLAFOKA 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 208 SSIYVKEGEQVEFSPFLAFVTEKLTGSGELMWQAEKSSKSWITFDLKNKEVSVKQVT 267
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SSTYVKEGEQVEFSPFLAFVTEKLTGSGELMWQAEKSSKSWITFDLKNKEVSVKQVT 240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 268 QDPKLOMGKRLPLHLTLPOALPOYAGSGNLTALAEKTKLHGEVNLVVMRATQLOKNT 327
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 QDPKLOMGKRLPLHLTLPOALPOYAGSGNLTALAEKTKLHGEVNLVVMRATQLOKNT 300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 328 CEVWGPTSPKMLSLKLENKKAIVSKREKPVWVNLNPEAGMOCCLSDSGQVLLSNIKVL 387
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 CEVWGPTSPKMLSLKLENKKAIVSKQAKAVWVNLNPEAGMOCCLSDSGQVLLSNIKVL 360
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 388 PTWSTPVP 396
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 PTWSTPVP 369
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 8

```

ID 088ZT8 PRELIMINARY; PRT; 457 AA.
AC 088ZT8;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Lymphocyte antigen CD4.
OS Calitrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrix.
OC NCBI_TaxID=9483;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22174698; PubMed=12186836;
RA Labonte J.A., Babcock G.J., Patel T., Sodroski J.;
RT "Blockade of HIV-1 infection of New World Monkey Cells Occurs
RT Primarily at the Stage of Virus Entry.";
RL J. Exp. Med. 196:431-445(2002).
DR EMBL: AF452616; AAN14532.1; -

```







RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Spleen;  
 RA Struhsberg R.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC053984; AAH53984.1; -  
 KW Hypothetical protein  
 SQ SEQUENCE 470 AA; 51204 MW; 778CF34521483F1A CRC64;

Query Match 37.4%; Score 1277; DB 4; Length 470;  
 Best Local Similarity 48.1%; Pred. No. 8.8e-89;  
 Matches 232; Conservative 33; Mismatches 104; Indels 178; Gaps 15;

QY	30	LGKGGTVELTCTAS--QKKSIOFHW-----KNSQITILNQSGFLLKGGPSK--LND 78
DB	30	LVOQGGSLRLSCVASFQTLNNYDMHWVRQIGKLEWVSKIGTAGDRVYAGVKGKFTIS 89
QY	79	RADSRSLWDQGNFLLIKNLKEDSDTYICEVEDQKEVQLLVFGLTANSDTHLLQGS 138
DB	90	RENADQSLYLQNM-----SLRVGDAAVYYC-----ARGGRMAPLGAFDIWGQ 133
QY	139	LTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWTCTVLQNGKVEFKID 198
DB	134	TWTVSSASTKGSPVFLPAPSSKSTSG--TALAGCL----- 168
QY	199	IVVLAFOKASIVYKKEGQVEFSPFLAFTVEKLTGSGELMWQAEKASSKSWITFDLKN 258
DB	169	-----VKDYPPEPYTVS-----WNSGALTSG----- 189
QY	259	KEVSVKRVTPQDPLQWKKPLHLTLPOALPOVAGSGNLTLALAEATGKHQEVNLVWR 318
DB	190	-----VH-TPAVL-QSSGIYSLSSVVTVPSSSLGTQTYI----- 222
QY	319	ATOLQKLTCEVWGPSTPKMLMLSLKLENKAKVSKKEKVVWVLPBAGMWQCLSDSGQV 378
DB	223	-----CNV-----NHR----- 229
QY	379	LIESNIKVLPTWSTPVEPKSCDKHTTCCPCAPABELLGGPSVFLPPPKKDTLMISRTPEV 438
DB	230	-----SNTKV-----DKKVEPKSCDKHTTCCPCAPABELLGGPSVFLPPPKKDTLMISRTPEV 282
QY	439	TCVVVDVSHEDPEVKENMYVDGVEVHNATKPREEQYNSTRVSVLTVLHODMLNGKEY 498
DB	283	TCVVVDVSHEDPEVKENMYVDGVEVHNATKPREEQYNSTRVSVLTVLHODMLNGKEY 342
QY	499	KCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAV 558
DB	343	KCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAV 402
QY	559	EWESNQGPNNTKTPPVLDSDGSFPLYSKLTVDKSRWQGNVFCSVNHEALHNHYTQK 618
DB	403	EWESNQGPNNTKTPPVLDSDGSFPLYSKLTVDKSRWQGNVFCSVNHEALHNHYTQK 462
QY	619	SLSLSPG 625
DB	463	SLSLSPG 469

RESULT 11  
 0727P5 PRELIMINARY; PRT; 469 AA.

AC 0727P5;  
 DT 01-OCT-2003 (Tremblrel. 25, Created)  
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Spleen;  
 RA MEDLINE=22388257; PubMed=12477932;  
 RA Struhsberg R.L.; Feingold E.A.; Grouse L.H.; Derge J.G.;  
 RA Klausner R.D.; Collins F.S.; Wagner L.; Shemen C.M.; Schuler G.D.;  
 RA Altschul S.F.; Zeeberg B.; Buetow K.H.; Schaefer C.F.; Bhat N.K.;  
 RA Hopkins R.F.; Jordan H.; Moore T.; Max S.I.; Wang J.; Hsieh F.;  
 RA Datchenko L.; Marusina K.; Farmer A.A.; Rubin G.M.; Hong L.;  
 RA Stapleton M.; Soares M.B.; Bonaldo M.F.; Casavant T.L.; Scheetz T.E.;  
 RA Brownstein M.J.; Usdin T.B.; Toshiyuki S.; Carninci P.; Frange C.;  
 RA Raha S.S.; Loquellano N.A.; Peters G.J.; Abramson R.D.; Mullahy S.J.;  
 RA Bosak S.A.; McEwan P.J.; McKernan K.J.; Malek J.A.; Gunaratne P.H.;  
 RA Richards S.; Worley K.C.; Hale S.; Garcia A.M.; Gay L.J.; Huiyk S.W.;  
 RA Villalón D.K.; Muzny D.M.; Sodergren E.J.; Lu X.; Gibbs R.A.;  
 RA Fahy J.; Helton E.; Kerteman M.; Madan A.; Rodriguez S.; Sanchez A.;  
 RA Whiting M.; Madan A.; Young A.C.; Shevchenko Y.; Bouffard G.G.;  
 RA Blakesley R.W.; Touchman J.W.; Green E.D.; Dickson M.C.;  
 RA Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.; Butlerfield Y.S.;  
 RA Krzywinski M.I.; Skalska U.; Smalins D.E.; Scherch A.; Schein J.E.;  
 RA Jones S.J.; Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Spleen;  
 RA Struhsberg R.;  
 RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC051328; AAH51328.1; -  
 KW Hypothetical protein  
 SQ SEQUENCE 469 AA; 51395 MW; C8D5BE12BAAF795C CRC64;

Query Match 37.1%; Score 1266; DB 4; Length 469;  
 Best Local Similarity 96.7%; Pred. No. 6.1e-88;  
 Matches 236; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

QY	382	SNIKVLPTWSTPVEPKSCDKHTTCCPCAPABELLGGPSVFLPPPKKDTLMISRTPEVTCV 441
DB	229	SNTKV-----DKKVEPKSCDKHTTCCPCAPABELLGGPSVFLPPPKKDTLMISRTPEVTCV 284
QY	442	VVDVSHEDPEVKENMYVDGVEVHNATKPREEQYNSTRVSVLTVLHODMLNGKEYCK 501
DB	285	VVDVSHEDPEVKENMYVDGVEVHNATKPREEQYNSTRVSVLTVLHODMLNGKEYCK 344
QY	502	VSNNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVE 561
DB	345	VSNNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVE 404
QY	562	SNQGPENNTKTPPVLDSDGSFPLYSKLTVDKSRWQGNVFCSVNHEALHNHYTQKSL 621
DB	405	SNQGPENNTKTPPVLDSDGSFPLYSKLTVDKSRWQGNVFCSVNHEALHNHYTQKSL 464
QY	622	LSPG 625
DB	465	LSPG 468

RESULT 12  
 08NF17 PRELIMINARY; PRT; 509 AA.

AC 08NF17;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE FLJ00385 protein (Fragment).  
 GN FLJ00385.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Spleen;



RA Jikuya H., Takano J., Kikuno R., Nagase T., Ohara O.;  
 RT "The nucleotide sequence of a long cDNA clone isolated from human  
 RT spleen."  
 RL Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AK090464; BAC03445.1; -.  
 DR PIR; A45874; A45874.  
 DR InterPro; IPR007110; Ig-1-like.  
 DR InterPro; IPR003597; Ig\_c1.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00047; Ig1.3.  
 DR SMART; SM00407; IGc1.3.  
 DR PROSITE; PS50835; IG\_LIKE.3.  
 DR PROSITE; PS00290; IG\_MHC.2.  
 DR NON\_TER  
 FT 1  
 SQ SEQUENCE 509 AA; 56111 MW; 089498076E863C CRC64;

Query Match 37.1%; Score 1265; DB 4; Length 509;  
 Best Local Similarity 91.7%; Pred. No. 8.1e-88;  
 Matches 231; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

OY 395 EPKSCDHTPCPCPAPAPLGGPSVFLPPPKDTLMISRTPEVTCVVDVSHEDPEYKF 454  
 DB 209 EPKSCDTPPCPCPCAPAPLGGPSVFLPPPKDTLMISRTPEVTCVVDVSHEDPEYQF 268  
 OY 455 NMVYDGVENNAKTPREEQYNSTYRVSVLTVLHQMNLNGEKYCKVSNKALPAPIEKT 514  
 DB 269 KMVYDGVENNAKTPREEQNFSTFRVSVLTVLHQMNLNGEKYCKVSNKALPAPIEKT 328  
 OY 515 ISKAKGPREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAVMEWESGGPENNYKTP 574  
 DB 329 ISKTKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAVMEWESGGPENNYKTP 388  
 OY 575 PVLDSGDSFFLYSKLTVDKSRMQGNVFSQVMEHALHNHYTQKSLSPGLQDETCAE 634  
 DB 389 PVLDSGDSFFLYSKLTVDKSRMQGNVFSQVMEHALHNHYTQKSLSPGLQDETCAE 448  
 OY 635 AODGELDLGKMTT 646  
 DB 449 AODGELDLGKMTT 460

RESULT 13  
 ID Q72351 PRELIMINARY; PRT; 482 AA.  
 AC Q72351;  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DB Hypothetical protein DKFZp686N02209.  
 GN DKFZp686N02209.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CX NCBT\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Human rectum tumor;  
 RA Bleecker H., Boecker M., Mewes H.W., Weil B., Amid C., Oeanger A.,  
 RA Fobo G., Han M., Wiemann S.;  
 RL Submitted (JUN-2003) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; BX518118; CAD98026.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 482 AA; 52852 MW; EDA75F1901D1A034 CRC64;

Query Match 37.0%; Score 1264.5; DB 4; Length 482;  
 Best Local Similarity 96.3%; Pred. No. 8.2e-88;  
 Matches 235; Conservative 0; Mismatches 5; Indels 4; Gaps 1;

OY 382 SNIKLPWSTPVEPKSCDHTPCPCPAPAPLGGPSVFLPPPKDTLMISRTPEVTCV 441  
 DB 242 SNTKV---DKVPEKSCDHTPCPCPAPAPLGGPSVFLPPPKDTLMISRTPEVTCV 297  
 OY 442 VVDVSHEDPEYKFMVYDGVENNAKTPREEQYNSTYRVSVLTVLHQMNLNGEKYCK 501

DB 298 VVDVSHEDPEYKFMVYDGVENNAKTPREEQYNSTYRVSVLTVLHQMNLNGEKYCK 357  
 OY 502 VSNKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAVME 561  
 DB 358 VSNKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAVME 417  
 OY 562 SNGPENNYKTPPVLDSGDSFFLYSKLTVDKSRMQGNVFSQVMEHALHNHYTQKSL 621  
 DB 418 SNGPENNYKTPPVLDSGDSFFLYSKLTVDKSRMQGNVFSQVMEHALHNHYTQKSL 477  
 OY 622 LSPG 625  
 DB 478 LSPG 481

RESULT 14  
 ID Q96P08 PRELIMINARY; PRT; 679 AA.  
 AC Q96P08;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DB Factor VII active site mutant Immunocjugate.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CX NCBT\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21477448; PubMed=11593034;  
 RA Hu Z., Garen A.;  
 RT "Targeting tissue factor on tumor vascular endothelial cells and tumor  
 RT cells for immunotherapy in mouse models of prostatic cancer."  
 RI Proc. Natl. Acad. Sci. U.S.A. 98:12180-12185(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Hu Z., Garen A.;  
 RL Submitted (FEB-2003) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF272774; AAK58686.2; -.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR00742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR001438; EGF\_11.  
 DR InterPro; IPR006209; EGF\_11.  
 DR InterPro; IPR002383; GLA\_blood.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; Ig-1-like.  
 DR InterPro; IPR003597; Ig\_c1.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR001254; peptidase\_S1.  
 DR InterPro; IPR001314; peptidase\_G1A.  
 DR InterPro; IPR000294; vitk\_dep\_G1A.  
 DR Pfam; PF00008; EGF.2.  
 DR Pfam; PF00594; gla.1.  
 DR Pfam; PF00047; Ig1.2.  
 DR Pfam; PF00089; trypsin.1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00010; EGFBLDOD.  
 DR PRINTS; PR00001; GLA\_BLOOD.  
 DR SMART; SM00181; EGF.2.  
 DR SMART; SM00179; EGF\_CA.1.  
 DR SMART; SM00069; GLA.1.  
 DR SMART; SM00407; IGc1.2.  
 DR SMART; SM00020; TRY\_SPC.1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL.1.  
 DR PROSITE; PS00022; EGF\_1.1.



DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS00011; GLUT CARBOXYLATION; 1.  
 DR PROSITE; PS00835; IG\_Like; 2.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_SER; 1.  
 DR PROSITE; PS00135; TRYPSIN\_HIS; 1.  
 SQ SEQUENCE 679 AA; 75552 MW; 0B0023AE70A067A1 CRC64;

Query Match 36.8%; Score 1258; DB 4; Length 679;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-87;  
 Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 395 EPKSCDKHTPCPCPAPELLGGPSVFLPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKF 454  
 DB 448 EPKSCDKHTPCPCPAPELLGGPSVFLPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKF 507  
 QY 455 NMVYDGEVHNAKTKREEQVNSTYRVSVLTITVLHQMILNGKEYCKVSNKALPAPIEK 514  
 DB 508 NMVYDGEVHNAKTKREEQVNSTYRVSVLTITVLHQMILNGKEYCKVSNKALPAPIEK 567  
 QY 515 ISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTP 574  
 DB 568 ISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTP 627  
 QY 575 PVLDDSGSFYKSKLTVDKSRWQQGNVSCSVMHEALHNHYTQKSLSLSPG 625  
 DB 628 PVLDDSGSFYKSKLTVDKSRWQQGNVSCSVMHEALHNHYTQKSLSLSPG 678

## RESULT 15

Q9XS78 PRELIMINARY; PRT; 455 AA.

AC 09XS78; 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE T-cell surface glycoprotein CD4.  
 OS Delphinapterus leucas (Beluga whale).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;  
 OC Monodontidae; Delphinapterus.  
 OX NCBI\_Taxid=9749;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Thymus;  
 RA MEDLINE=9216435; PubMed=10199913;  
 RA Romano T.A., Ridgway S.H., Felten D.L., Quaranta V.;  
 RT "Molecular cloning and characterization of CD4 in an aquatic mammal,  
 RT the white whale Delphinapterus leucas.";  
 RL Immunogenetics 49:376-383(1999).  
 DR EMBL; AF071799; IAD23738.1; -.  
 DR HSSP; P01730; IWIQ.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:006955; P:immune response; IEA.  
 DR InterPro; IPR000973; CD4\_TcRg.  
 DR InterPro; IPR007110; IG\_Like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; Ig\_3.  
 DR PRINTS; PR00692; CD4TCANTIGEN.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG\_Like; 2.  
 SQ SEQUENCE 455 AA; 50499 MW; AA532FD4411AA5D1 CRC64;

Query Match 34.7%; Score 1185.5; DB 6; Length 455;  
 Best Local Similarity 60.2%; Pred. No. 7.8e-82;  
 Matches 254; Conservative 46; Mismatches 107; Indels 15; Gaps 5;

QY 1 NMRGVPFRLLLVLTALIPATQGNKVLGKGGDTVELTCTASQKSIQFHKNKSNQIK 60  
 DB 1 MDPRTSLRHLFLVLQVLMIPAGTQGGKVVILGKAGELAIELPKASQNKSLFFSMKNSYQTK 60

QY 61 ILNGSFLTKGSPKTLNDRADSRSLMDQGNFPLIKKLIKIEDSDTYICEVEDQKEEYOL 120  
 DB 61 ILGRIGYFMHKGASLHRSVESKINLMDQGSFPLVIKOLEVDSGTIICEVEDKKEIYEL 120  
 QY 121 LVFGLTANSDFHLLQGSLTTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDPSG 180  
 DB 121 QVFRLLTASSDFRLLLGSLTTLTLEPGSGNSPSVQKQKGNKRNKAKSLSPQVQLQDSG 180  
 QY 181 TWTCVTLNQKKEKVEKIDIVLAPOKASSIYKKEGEQVEPFLAFVTEKLTGSGEL-W 239  
 DB 181 TWTCVTSQAQQTLLVFNKHIILVLAFOGVSSYVAKGEGEMNFSFPLTFDENL--SGELSW 238  
 QY 240 WQAEASASSKSWITFDLKNKEYSVRYTPDPLQMGKLLPHLTLPQALPOYAGSGNLT 299  
 DB 229 LQAKNNSPESVITTKLNGKVTYVAKARDLKLRSKALPLHITLPPQALPOYAGSGNLT 298  
 QY 300 ALEATGKLHDEVNILVNRATQQLNLTCEVWGPTSPKMLSLKENKARKVSKKEKPVW 359  
 DB 299 NL--TKGKLYGEVNLVNRVTKSPNSLTCEVLTGPTSPRLISLKKENGSMRVSDDQKLV 356  
 QY 360 VLNPAAGMWQCCLSDSGQVLTLESNITKVLPTMSTPYEPKSCDKHTPCPCPAPELLGGPSV 419  
 DB 357 VLGPEAGMWQCCLSDSGKVLLESKYKILP-----LV-----LAHAMPKLLAVLGGITSL 406  
 QY 420 FL 421  
 DB 407 LL 408

## RESULT 16

Q86TT2 PRELIMINARY; PRT; 354 AA.

AC 086TT2; 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Human full-length cDNA clone CS0D1019F20 of placenta of Homo sapiens  
 DE (Human) (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Genoscope;  
 RL Submitted (FEB-2003) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Li W.B., Gruber C., Jessee J., Polayes D.;  
 RT "Full-length cDNA libraries and normalization."  
 RT Submitted (FEB-2003) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; BX248278; CAD62606.1; -.  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 DR InterPro; IPR007110; IG\_Like.  
 DR InterPro; IPR003597; IG-cl.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR Pfam; PF00047; Ig\_1.  
 DR SMART; SM00407; IGc1; 3.  
 DR PROSITE; PS00835; IG\_Like; 3.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KM Plasmid.  
 FT NON\_TER.  
 SQ SEQUENCE 354 AA; 39125 MW; 23B80BF4D2B87A92 CRC64;

Query Match 34.3%; Score 1171; DB 4; Length 354;  
 Best Local Similarity 92.6%; Pred. No. 7e-81;  
 Matches 214; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 395 EPKSCDKHTPCPCPAPELLGGPSVFLPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKF 454  
 DB 123 EPKSCDKHTPCPCPAPELLGGPSVFLPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKF 182







DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
 DE CD4 antigen precursor.  
 OS Felis silvestris catus (Cat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 NCBI\_TaxID=9685;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Miyazawa T.;  
 RN Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9216516; PubMed=1537604;  
 RA Noriame J., Miyazawa T., Kawaguchi Y., Tohya Y., Kai C., Mikami T.;  
 RT "A CDNA encoding feline CD4 has a unique repeat sequence downstream of  
 the V-like region."  
 RL Immunology 75:74-79(1992).  
 DR EMBL; AB000483; BAA19124.1; -.  
 DR HSSP; P01730; IMIQ.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0006955; P:immune response; IEA.  
 DR InterPro; IPR000973; CD4\_TcAg.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; Ig\_3.  
 DR PRINTS; PR00692; CD4TCANTIGEN.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 2.  
 FT SIGNAL.  
 FT CHAIN 1 26 POTENTIAL.  
 FT CHAIN 27 474 POTENTIAL.  
 SQ SEQUENCE 474 AA; 52243 MW; D946DD4BEAD00EC CRC64;  
 Query Match 31.8%; Score 1087; DB 6; Length 474;  
 Best Local Similarity 57.6%; Pred. No. 2.6e-74;  
 Matches 239; Conservative 49; Mismatches 99; Indels 28; Gaps 6;  
 QY 1 NNRGVPRRLILVLTALPAAT-OGNKVILGKGGDTVELTCTASOKKSIOFHMKNNOI 59  
 DB 1 NMQGAVFRLLLVLTQVLMKAAVPOGKEVVLGAGTABELPCQASQKKTMTTMRSLSSGV 60  
 QY 60 KILGNQGSFL-TKGSPLKNDRASSRLMDQGNFPLIINKLKIEDSDTYICEVEDQKEV 118  
 DB 61 KILSGHSLCLTGSSKLTREPSKILMDQGSFLVILKSLQVADSIGITCEVENKKEV 120  
 QY 119 QLVVFGLTANSD-----THLLQGQSLTTLTLESPPGSSPVQCSPPK 161  
 DB 121 ELTVFGLTAKVDPSSGSGSSSSSTSTSTSTSYLLQGQSLTTLTLESPPSSNPVSQVMKGP 180  
 QY 162 NIQGGKTLTSSQLELQDSTGWTCTVLQNKVYFKIDIVLAFOKASSIVYKKEGOVEF 221  
 DB 181 SSGVSHSLSLQLEQSGTCTCTVSSQSKTVFNTNLTIVLAFRYVSNIVYAKKEGOVEF 240  
 QY 222 SFLPAFTVEKLTGSGELMWQAEASSSKSWITFDLKNKEVSVKRYTODPKLQMGKPLH 281  
 DB 241 SFLPLAFEDENLGN--LRWKAEGAPSSLLMISFTLKNQSLVKEVDPSYKLOMMSLPR 298  
 QY 282 LTLPAALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQLOKNTTCEYWGFTSPKMLLS 341  
 DB 289 FLPLPVLSRYAGSNLTLLVD--KQLOQEVKLVMRATQSGNNLTCEVLTGPTSDLTLS 356  
 QY 342 LKLEKKEAVSKREKRVVWLNPEAGMOCILSDSGVILLESNIKVL-----TWS 391  
 DB 357 LKLKGOAAKVSQKQKRVVEDAEAGTWOCLLSHKDKVLLASFAEVLPPVLTTRTWT 411

DE T-cell differentiation antigen.  
 GN CD4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8804159; PubMed=2823269;  
 RA Gorman S.D., Tourville B., Farnes J.R.;  
 RT "Structure of the mouse gene encoding CD4 and an unusual transcript in  
 brain."  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).  
 DR EMBL; M17080; AAA37402.1; -.  
 DR EMBL; M17077; AAA37402.1; JOINED.  
 DR EMBL; M17078; AAA37402.1; JOINED.  
 DR EMBL; M17079; AAA37402.1; JOINED.  
 DR HSSP; P01730; IMBR.  
 DR MGD; MGI:88335; CD4.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0006955; P:immune response; IEA.  
 DR InterPro; IPR000973; CD4\_TcAg.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; Ig\_2.  
 DR PRINTS; PR00692; CD4TCANTIGEN.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 SQ SEQUENCE 457 AA; 51368 MW; 24AB19EDA285BD0 CRC64;  
 Query Match 28.9%; Score 985; DB 11; Length 457;  
 Best Local Similarity 53.6%; Pred. No. 1.5e-66;  
 Matches 210; Conservative 62; Mismatches 110; Indels 10; Gaps 7;  
 QY 1 NNRGVPRFH-LTLVLTALPAATQGNKVILGKGGDTVELTCTASOKKSIOFHMKNNOI 59  
 DB 1 MCRALSRLRLILVLTQSLQVLAIVQEKTLVIGKEGSAELCESQKKTITVFTWKFSDR 60  
 QY 60 KILGNQGSFLTKG--PSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKE 116  
 DB 61 KILGKHGKGVILRGSSPSQF-DREPSKKGAWEKSFPLIINKLKIEDSDTYICELENKE 119  
 QY 117 EVQLVPELTPANSDPHLQGGSLTTLTLES-PPGSSPVQCSPPKRNIOGGKTLSSQLE 175  
 DB 120 EVELVWFVETSPGSLTLQGGSLTTLTLDNSKVSNPTECHKKGKGVSGSKVLSMSNLR 179  
 QY 176 LQDSTGWTCTVLQNKVYFKIDIVLAFOKASSIVYKKEGOVEFSPFLAFTVEKLTGS 235  
 DB 180 VQDSFPMWCTYTLQDKKRWFGMTLSVLGFQSTAITAYKSEBSAEFSFLNFAEE--NGW 237  
 QY 236 GELMWQAEASSSKSWITFDLKNKEVSVKRYTODPKLQMGKPLHLLTLPAALPOYAGSG 295  
 DB 238 GELMWKAKDGFPGQWISFISIKNKESVQKSTKDLQKLTPLTKIPQVSIQFAGSG 297  
 QY 296 NLTALAEKTKLHQEVNLVVMRATQLOKNTTCEYWGFTSPKMLLSLKLEKKEAVSKRE 355  
 DB 298 NLTLLVD--KQTLHQEVNLVVMKVAQALNNTLTCEYWGFTSPKMLLTQEQNEARVSEQ 355  
 QY 356 KPVWVLPNPEAGMOCILSDSGVILLESNIKVL 387  
 DB 356 KVVQVVAPEETGLMOCCLSEGDVKMDSRIQVL 387

RESULT 20  
 061396 PRELIMINARY; PRT; 457 AA.  
 AC 061396;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

RESULT 21  
 055054 PRELIMINARY; PRT; 433 AA.  
 AC 055054;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE T4 surface glycoprotein (Fragment).



```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86097446; PubMed=3501122;
RA Maddon P.J., Molineux S.M., Maddon D.E., Zimmerman K.A., Godfrey M.,
RA Alt F.W., Chess L., Axel R.;
RT "Structure and expression of the human and mouse T4 genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:9155-9159(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Maddon P.J., Molineux S.M., Maddon D.E., Zimmerman K.A., Godfrey M.,
RA Alt F.W., Chess L., Axel R.;
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF045882; AAC01764.1; -.
DR HSSP; P01730; 1MR.
DR MGD; MGI:88335; Cd4.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR006973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PRO0692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER
FT TER
FT SEQUENCE 433 AA; 48590 MW; AB19330750A8499A CRC64;
SO SEQUENCE 433 AA; 48590 MW; AB19330750A8499A CRC64;

Query Match 28.0%; Score 957; DB 11; Length 433;
Best Local Similarity 51.0%; Pred. No. 1,8e-64;
Matches 205; Conservative 61; Mismatches 114; Indels 22; Gaps 7;

QY 24 QGNKRVLGKGDYELTCTASQKSIQFMKNSQIKILGNQ--SFLTK--PSKLNDR 80
DB 1 QGKTLVLGKEGESALPCSSQKKTIVPTWKFSDDRKILGQHGKVLIRGSPSQ--DRF 59
QY 81 DSRSLMNGQNPPLIKLKLTEDSTTYCEVEDQKEVQLLVFGITANSDFHLLQGGSLT 140
DB 60 DSKGAMKEGSPPLINKLKMEDSQTYCELENKEEVELWVFATFSPGSLIQGOSLT 119
QY 141 LTLRS-PPGSSPVSQCRPRGNLQGGKTLVSQLELDSSGNWCTTVLQONOKVEFKDI 199
DB 120 LTLDSNKNVSNPLTECKRKKGKVVSGKVLMSNLRVDSDFMNCVTYLDQKMFQKTL 179
QY 200 VLAIFQKASSIVYKKEGQVEFSPFLAFTVEKLTGSGELMWQAERASSSSKMITFDLKNK 259
DB 180 SVLGQSTALTAAYKSEGSASFSPFLNPAEE--NMGELMWKAEKDSFPQWISPSIKNK 237
QY 260 EVSVKRVYQDRLQMGKLLPHLTLPOLPYAGSGNLTLLBAKTLGHQEVNLVYVRA 319
DB 238 EVSVQKSTKDLTKLTKETLPLTKIPQVSLQFAGSGNLTLLD--KGTLHQEVNLVYVRA 295
QY 320 TOLQNLNLCCEWGPSPKLTMLSLKLENKAKVSKREKFWYVLANPEAGMWOCILDSGQVL 379
DB 296 AQLNNTLTCEWGPSPKRLTLKQENDEKAVSEKQKVVQVAPETGLMOCLISBGDKVK 355
QY 380 LESNIKVLPTWSTVPEKSCDKTHTCPAPAPELLGSPVFL 421
DB 356 MDSRIQVLSRGVNVFLAC-----VLGGSFATL 384

RESULT 22
Q95M34 PRELIMINARY; PRT; 337 AA.
AC Q95M34;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Immunoglobulin gamma 1 heavy chain constant region (Fragment).

```

```

GN IGHCL.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Wagner B.;
RA Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98383416; PubMed=9717671;
RA Wagner B., Overesch G., Sheoran A., Holmes M., Richards C.,
RA Leibold W., Radbruch A.;
RT "Organization of the equine immunoglobulin heavy chain constant region
genes. III. Alignment of c-mu, c-gamma, c-epsilon and c-alpha genes.";
RL Immunobiology 159:105-119(1998).
DR EMBL; AJ30675; CAC44624.1; -.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_2.
DR SMART; SM00407; IGV; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
FT NON_TER
FT TER
FT SEQUENCE 337 AA; 37438 MW; A60BF2B01DEFD1F6 CRC64;
SO SEQUENCE 337 AA; 37438 MW; A60BF2B01DEFD1F6 CRC64;

Query Match 26.9%; Score 919; DB 6; Length 337;
Best Local Similarity 61.3%; Pred. No. 1e-61;
Matches 173; Conservative 38; Mismatches 43; Indels 28; Gaps 5;

QY 371 LLSDDGVLLESNIKV-LPTWSTP-----VEP-----KSCDKTHTC 405
DB 56 VLQSSGYSLSNWTVPASTWTSETYICNVVHAASNKVDKRIPIIDNNQKVDMS-KC 114
QY 406 PCPAPPELLGSPVFPFPKCDTLMISRTPEVTCVVDVSHDEPEKVMYVDGVEVHN 465
DB 115 PCPAPPELLGSPVFPFPKCDTLMITRPEVTCVVDVSGENPDVCFMVMYDGVVVRT 174
QY 466 AKTKPREQVNSTYRVSVUTVLHOMLNKEKCKKSNKALPAPIEKTISKAGQREP 525
DB 175 ATTRPKKEQNSTYRVSVVTLRIQHDVLSGERKCKKNNALPPIERTTKTKGRQEP 234
QY 526 QVYTLPSRDELTFKQVSLCLVKGFPYSDIAVEMESNGQP--ENNYKTTIPVLDSDGSF 583
DB 235 QVYTLAPHPBELSKSKSVCLVKDFFPPEINIEQNSGPELETKSTTQAQDSGDSY 294
QY 584 FLVSKLTVDKSRWQQGNVFSQVMEHALNHNHTQKSLSPG 625
DB 295 FLVSKLSDVDRNRWQGTTFTCGVWHEALNHNHTQKNVSKNPG 336

RESULT 23
Q7TMK1 PRELIMINARY; PRT; 470 AA.
AC Q7TMK1;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CZECH 11; TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,

```



RA	Diacchenko L., Martusina K., Farmer A.A., Rubin G.M., Hong L.
RA	Stapleton M., Soares M.B., Bonaldi M.F., Casaravito T.L., Sheetz T.E.,
RA	Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA	Boeak S.A., McEwen P.J., McKernan K.T., Malek J.A., Gunaratne P.H.,
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA	Fahay J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA	Rodriguez A.C., Gilwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA	Krzyszinski M.I., Skalska U., Smalhus D.E., Scherch A., Schein J.E.,
RA	Jones S.J., Marra M.A.;
RT	"Generation and initial analysis of more than 15,000 full-length human
RT	and mouse cDNA sequences."
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN	(2)
RP	SEQUENCE FROM N.A.
RC	STRAIN-CZECH II; TISSUE-Breast tumor;
RA	Strausberg R.;
RL	Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
DR	EMBL; BC055910; AAHS5910.1; -.
SK	Hypothetical protein.
SQ	SEQUENCE 470 AA; 51727 MW; 6D90A8DF996BB090 CRC64;
Query Match	25.1%; Score 857; DB 11; Length 470;
Best Local Similarity	39.1%; Pred. No. 8,6e-57;
Matches 205; Conservative	58; Mismatches 129; Indels 132; Gaps 15,
OY	147 PGSSPVGCRSPRGKNIQG-----GKLTSVQLDLSDSGTWTCVLONOKKVEFK 196
Db	33 PGASVKISCKA-SGYTFGGYWMWKSHGKLEMGIVLPNSGDTG---YNQR--FK 84
OY	197 IDIVLAQKASSIYYKKGEQEVSFPLAFVEKLGTSGELMWAQEAASSKSMTFDL 256
Db	85 -GKALITVDKSSSTAYME-----LNSLTSEDAVVYCAITYYSGSITWYFDV 129
OY	257 --KNKEVSVKRVTODPKLOMGKKPLPHILTLPGALPOVAGS-----NLTALAIAK---- 304
Db	130 WGAQTIVTVSSAT-----TTAPSVPLVPGCGDTSGSSVTLGLVKGPFPE 175
OY	305 -----TGKLGHEVNLY-----VMRAITOLQKULTCEVMGPRTSKMLLS 341
Db	176 PVTKAMNYGALLSGGRVTSSYLQSGFSYSLSLVTPSSITWPSQTVICVAHPAS----- 229
OY	342 LKLENKAQKAVSKREKPVWVWLNPFAQMOCCLSDSGQVLLSNIKYLPMTSPFPVEPKSGDK 401
Db	230 -----KTELIRKRIIP-----RIPKSTIP--PQS--- 250
OY	402 THTCPBPAPELLGGPSVFLLPPPKPKDTLMISRTEVTCVVVDVSHEDPEVKFMNVYDGV 461
Db	251 -----SCPFGNITLGGPSVFIFPPPKKDALMISLTFRKVCITCVVVDVEDDDPDVHVSFFVDNK 305
OY	462 EVNAHKAKTPREQQNSTRVRSVLTVLADODLNAGEYKCYKSNKALPAPIETKTISKAKGQ 521
Db	306 EVHTAMTPTPRAEQNSTFRVRSALPIQHODMWKGEKFCCKVNNKALPAPIETKITSKPKGR 365
OY	522 PREPOVYTLPSPRDELTKNOVSLTGLVGFYPDSIDAVEWSNGOEENNYKATTPTPLDSDG 581
Db	366 AQTPQVYTIIPPREQMSKKKQSLSLTGLTNFSPSEALSVERNGELEGEOYKATPTPLDSDG 425
OY	582 SPFLYSKLTVDKSRMQGNVFSQSYVMHALNHNTQKSLSPG 625
Db	426 TYFLYSKLTVDPTDSDLQGEIFTCSSVYHEALHNHNTQKSLSPG 469
RESULT 24	
O9RIA4	
ID O9RIA4	PRELIMINARY; PRT; 437 AA.
AC O9RIA4	
DT 01-MAY-2000 (TREMBLrel. 13, Created)	
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)	

DE Gammal heavy chain of Mab7 (Fragment).  
GN IGH-4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wilder K.G., Yu X., Ekramoddoullah A.K.M., Misra S.;  
RT "Cloning of cDNAs encoding for anti-white pine blister rust monoclonal  
RT antibody (Mab 7, its light and heavy chains) and construction of a  
RT single chain antibody (scFv)".  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF152372; AAD40243.1; .  
DR PIR; B45837; B45837.  
DR PDB; 1COK; 11-SEP-99.  
DR PDB; 1I91; 25-DEC-02.  
DR PDB; 1KCU; 11-MAY-02.  
DR MCD; MGI:96446; Igh-4.  
DR InterPro; IPR007110; Igh-like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig; 3.  
DR SMART; SM00406; Igv; 1.  
DR PROSITE; PS50835; Ig\_LIKE; 4.  
DR PROSITE; PS00290; Ig\_MHC; 1.  
FT NON\_TER 1 1  
FT TER 437 437  
SQ SEQUENCE 437 AA; 48142 MW; 5C3A7BB3BE7D697C CRC64;  
  
Query Match 24.9%; Score 850.5; DB 11; Length 437;  
Best Local Similarity 33.4%; Pred. No. 2,4e-56;  
Matches 202; Conservative 74; Mismatches 142; Indels 187; Gaps 15;  
  
QY 30 LGKKGDVVELTCTAS--QKSIQPHMKNSNQIKILNQGSFTLTGPKLNDRAASRLSM 87  
DB 10 LVKGGSLKLKSCAASGFTFSYANSWRQRPPEKLEWASSSSGIIYYTDSVGRFTY 69  
QY 88 ---DQGNFPLILKILKTIEDSDPTVCEVDEQKEEVQLVFGILTANSDPHLLQGSGLTLE 144  
DB 70 KDKORNLISLQMSLRBEDTMYTC-----AGDYSAWGPGLTVGS 112  
QY 145 SPSSSPSSVQCRSPRGKNIQGGKTLVSQLELQDSGTWTCTVLQNKKEFKIDIVLAF 204  
DB 113 AAKTTPSYVPLA-----GSAQTNSMTL----- 138  
QY 205 QKASSIYYKKEGEVEYFSPLATVYEKLTGSGELMQQAEARSSSKSMITPDLKKEVSK 264  
DB 139 -----GCLVKKYFPEPVT-----NNSGISLSC----- 162  
QY 265 RVTQDPKLQMGKPLHLTLTQALPQ--YAGSGNLTLALEAKTGKLGQEVNLVYMRATQL 322  
DB 163 -----VA-TTPAVLQSLYTLSSSVT-----VPSSTWP 189  
QY 323 QKQLTCEWGPSTSPKMLSLKENKEAKVSRREKPVVNLPEAGMOCLLSDSGVLLS 382  
DB 190 SEYTCNVAVHPAS-----SKKVDK----- 209  
QY 383 NIKVLPTWSTVPEPKSCDKTHTCP--CPAPELLGSPVPLPPPKPKDTLMISRTPEYTC 440  
DB 210 -----IYPRDGG--CKPCTCTVPEV--SSVFPPKPKDVLITLTPEKVT 251  
QY 441 VVVDVSHEDPEVKKMNVVDGVEVNNATKREBOYNTSYVSVLTVLHODMNGKRYKC 500  
DB 252 VVVDISKQDPEVQSMFTVDVEVHTAQTQPREDFNSTFRSSVSLPIHODMLNGKKEKC 311  
QY 501 KVSNAKLPALIEKTIKSAKQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEW 560  
DB 312 RVNSAAPAPRIEKTISTKGRPKAPQYITIPPRREGMAKQKVSGLTCMITDFFPEDIVEM 371  
QY 561 ESNQGPENNTKTPPVVLDSQGSFFLYSKLTVDKSRWQOQGVFSGVNHGALAHNYTKSL 620  
DB 372 QMNGQPEANNTKTPPVIMDTQGSYFVSKLVNQKSNMDEAGNTPTCSVLHBEGLAHNHTEKNL 431



OY 621 SLSPG 625  
DB 432 SHSPG 436

## RESULT 25

O8R3V9 PRELIMINARY; PRT; 469 AA.

ID O8R3V9  
AC O8R3V9;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN IGH-4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strauberg R.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC024405; AAH24405.1; -.  
DR PIR; B45837; B45837.  
DR MCD; MGI:96446; Igh-4.  
DR InterPro; IPR007110; Igh-1like.  
DR InterPro; IPR003006; Igh\_MHC.  
DR InterPro; IPR003596; Igh\_v.  
DR Pfam; PF00047; Igh; 3.  
DR SMART; SM00406; Igv; 1.  
DR PROSITE; PS50835; Igh\_LIKE; 4.  
DR PROSITE; PS00290; Igh\_MHC; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 469 AA; 51976 MW; 534793F155D05457 CRC64;

Query Match 24.9%; Score 850.5; DB 11; Length 469;  
Best Local Similarity 34.9%; Pred. No. 2.7e-56;  
Matches 189; Conservative 75; Mismatches 109; Indels 169; Gaps 15;

OY 147 PGSSPSVOC-----RSPRGK-----NIQGKTL5 170  
DB 33 PGSLRLSCAASGFTDYWSWVQPPGKALEWLGIRKANGYTTYASVSKGRFTIS 92  
OY 171 -----VSOLELQDSGTCTVQONQKVEFKIDIVLAFOKASIVYKKEGEQV 219  
DB 93 RDNOSITLYLQNMALRAEDSATYCA---RDRRSY-----YSGTSFAYMGQGLV 141  
OY 220 EFS-----FPIAFIVEKLTGSGELMWQAEKSSSSKSWITFDLKNKEVSVKRYTQDP 270  
DB 142 TVSAKTTTPPSVYPLA-----PGS-----AAQTNMVTILGCLVKGY----- 177  
OY 211 KLQGMKCLPLHLTPQALPOYAGSGNLTALAEAKTGKLGHOEV-----NLVYMRATOLQKN 325  
DB 178 -----FPEPVTVTWNSGSLSSGVHTFPAVLQSDLYTLSSSVTPVSSWTWPSQT 224  
OY 326 LTCEVWGPSTPKMLSLKLENKEAKVSKREKRVWVLANPEAGMWOCILSDSGVLLSNIK 385  
DB 225 VTCVNAHNAS-----STKVDK----- 241  
OY 386 VLPWSTVEPSPKCDKHTCP--CPABELLGGPSVFLFPKPKDTLMISRTPEVTCVV 443  
DB 242 -----IVPRDGG-----CKPCICTVPEV---SSVFIFPPKPKDVLITLTLPKVTQVV 286  
OY 444 DVSHEDPEVKFNWYDGVENNAKTKPREEQVNSTYRVSVLTVLHQMVLNGKEKCVS 503  
DB 287 DISKDDPEVQSFWDVDEVHTAQTKPREEQFNSTFRVSSELPIHQMLNGKEKCRVN 346  
OY 504 NKALPAPLEKTIISKAKQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPYSDIAVEMESN 563  
DB 347 SAAPFAPIEKTISKTKGRPKAPQVYTIIPPKQGMKDKVSLTCLMTDFPEDDIVEMQNN 406  
OY 564 GQPENNYKTTPPVLDSGSPFLYSKLTVDKSRMOGQNVFSCSVMEALHNHYTKSLSL5 623

DB 407 GQPAENYKNTQPIMDTDGSYFYVYSKLVQKSNWEAGNTFTCSVLHBEGLHNHHEKSLSHS 466  
OY 624 PG 625  
DB 467 PG 468

## RESULT 26

O99LC4 PRELIMINARY; PRT; 463 AA.

ID O99LC4  
AC O99LC4;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Similar to RIKEN cDNA 1810060009 gene.  
GN IGH-4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strauberg R.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC003435; AAH03435.1; -.  
DR PIR; B45837; B45837.  
DR HSSP; P01842; 7PAB.  
DR MCD; MGI:96446; Igh-4.  
DR InterPro; IPR007110; Igh-1like.  
DR InterPro; IPR003006; Igh\_MHC.  
DR InterPro; IPR003596; Igh\_v.  
DR Pfam; PF00047; Igh; 3.  
DR SMART; SM00406; Igv; 1.  
DR PROSITE; PS50835; Igh\_LIKE; 4.  
DR PROSITE; PS00290; Igh\_MHC; 1.  
SQ SEQUENCE 463 AA; 51007 MW; EAA674C6B8C30783 CRC64;

Query Match 24.7%; Score 844; DB 11; Length 463;  
Best Local Similarity 36.7%; Pred. No. 8.2e-56;  
Matches 186; Conservative 72; Mismatches 120; Indels 132; Gaps 14;

OY 136 GQSLITLLESPSPSPSVQCRSPRGK-NIQGKTL5-----VSOLELQDSGTCTVQLQN 189  
DB 61 GQGLEWGEIYIPSGSGNTYSEKFKGKATLTDDSSSTAYVHMLSLTSEDAVAFCA----- 116  
OY 190 QKVEFKIDIVLAFOKASIVYKKEGEQVFS-----FPIAFIVEKLTGSGELMW 240  
DB 117 -RSSYISYD-----FAYMGQTLVTVSAKTTTPPSVYPLA-----PGS----- 154  
OY 241 QAERASSSKSWITFDLKNKEVSVKRYTQDPKLGKCLPLHLTPQALPOYAGSGNLTALA 300  
DB 155 -----AAQTNMVTILGCLVKGY-----FPEPVTVTWNSGSLSSG 186  
OY 301 LEAKTGKLGHOEV-----NLVYMRATOLQKNLTCEVWGPSTPKMLSLKLENKEAKVSKRE 355  
DB 189 VHTFPAVLQSDLYTLSSSVTPSSTWPSFTVTCVNAHPAS-----STKVDKK- 235  
OY 356 KPVWVLANPEAGMWOCILSDSGVLLSNIKVLPTWSPVPEPKSGDKTHTCP--CPABEL 413  
DB 236 -----IVPRDGG-----CKPCICTVPEV---SSVFIFPPKPKDVLITLTLPKVTQVV 286  
OY 414 LGGSVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYDGVENNAKTKPREE 473  
DB 254 -----SSVFIFPPKPKDVLITLTLPKVTQVVVDISKDDPEVQSFWDVDEVHTAQTKPREE 310  
OY 474 QYNSTYRVSVLTVLHQMVLNGKEKCVSNKALPAPIEKTISKAKQPREPOVYTLPPS 533  
DB 311 QFNSTFRVSSELPIHQMLNGKEKCRVNSAAPFAPIEKTISKTKGRPKAPQVYTIIPP 370  
OY 534 RDELTKNOVSLTCLVKGFPYSDIAVEMESNGQPENNYKTTPPVLDSGSPFLYSKLVYDK 593  
DB 371 KEQMAKDKVSLTCLMTDFPEDDIVEMQNGQPAENYKNTQPIMDTDGSYFYYSKLVNQK 430



QY	SVWQOQNVSSCVMEHFAALNHNTOKSLSPG 625
DB	431 SNMEAGNTFTTCSVLHGLHNNHTKSLSPG 462
RESULT 27	
Q9DBL4	PRELIMINARY; PRT; 473 AA.
ID	Q9DBL4
AC	Q9DBL4;
DT	01-JUN-2001 (TREMBLrel. 17, Created)
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE	01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE	181006009RIK protein.
CN	IGH-1 OR 181006009RIK.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutelesta; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxId=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=Pancreas;
RX	MEDLINE=21085660; PubMed=11217851;
RA	Kawaji J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA	Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA	Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Aeshbunter M., Batalov S., Casavant T.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA	Schmitt L.M., Stambli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA	Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.O., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Guerninich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H.,
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA	Wynhauer-Botis A., Yoshida K., Hasegawa Y., Kawaji J., Kohseki S.,
RA	Hayashizaki Y.;
RT	"Functional annotation of a full-length mouse cDNA collection.";
RL	Nature 409:685-690(2001).
DR	EMBL; AK007918; BAB25349.1; -
DR	PIR; S26746; S26746.
DR	HSSP; F01842; 7FAB.
DR	MGI; MGI:96443; Igh-1.
DR	InterPro; IPR007110; Ig-1like.
DR	InterPro; IPR003006; Ig_MHC.
DR	InterPro; IPR003596; Ig_V.
DR	Pfam; PF00047; Ig_3.
DR	SMART; SM00406; Igv_1.
DR	PROSITE; PS50835; IG_LIKE; 4.
DR	PROSITE; PS00290; IG_MHC; 1.
SQ	SEQUENCE 473 AA; 51699 MW; 9DED57A514475FBB CRC64;
Query Match	24.6%; Score 840.5; DB 11; Length 473;
Best Local Similarity	34.2%; Pred. No. 1.6e-55;
Matches 208; Conservative 75; Mismatches 147; Indels 179; Gaps 17	
QY	30 LGKKDDYELTETTAQKSIQPHMKNSNDIKILGNQ-SFLTK-----GPSKLNDRADR 83
DB	30 LVPKASVYKISCKAGYTTTDY---INWVKRPPQGLEWIKIGPGSSITYNEKFKK 86
QY	84 RSLW---DQGNPLIINKLIKEDSPYICEVEDQKEVOGLVFGILTANS-----DTHLGG 136
DB	87 ATLTKADKSSSTAYMOLSLTSEDSAAVFC-----ARGDYDWPAYWG 129
QY	137 OSULTLLESPESSSVGCRSGRKNIGOGKTLVSQLELDPSGTMTCTVLONOKKVERK 196
DB	130 OCTLVVSAKTTAASVYPLAP-----VCCGTT----- 157
QY	197 IDIVIAFOKASSIYVKEGDEVERFPLAFVVEKLTGSGELMWOAERASSKSWITFDL 256

Dd	158	-GSSVTL--GCILVKYFP-----EPVT---LTWNSGLSSG-----	187
Oy	257	KNKEVSQKRVTPQPKLQMGKKLP.LHLTLP.OALPQYAGSGNTLLALEAKTKLHQENVLV	316
Dd	188	-----VA-TTPALLQ-----SGLYTLS-----SSVTY	208
Oy	317	MRAIOTOLKULTCFVGWPSTPKMLMSLKLENKAQYSKRKEPVWLINPEAGMOCILLDSG	376
Dd	209	TSMTSPSQITTCVNAHPAS-----STKVDKKIETP-----	237
Oy	377	QVLLESNIKLPFWMSIPVPKSCDXTHTCP.PAPPELLGGSVFLPPPKPYDTLMISRTF	433
Dd	238	-----RVPITONPCPLKECP.PCAAPDLIGGSVFIFPFXIKDVIIMISLP	283
Oy	437	EATCVVVVDVSHEDPEKFNFMYDVGYEVHNAKTKPREEOYNSTYVWSVLFVLHODMTLNGK	496
Dd	284	MATCVAVDVSEDDPDQAISM.FNVANVEVHTAQITREIDYSTLTLAVSA.LPIQHODMMSGK	343
Oy	497	EYACKVSNKALPAPIEKTIISKAKGP.PREPQVYTL.PPSRDELTKNOVSLTCLVKGFPYSDI	556
Dd	344	EFCKCKNNNALPPIEKTISKRGVPRAP.QVYVLP.PPAEMTKXKEFSLTCTMTGF.LPAEI	403
Oy	557	AZWESNGCPENNYKTTTPVLDSDGSFFLYSKLT.YDTSRKMOGVNFCSSVMHEALHHYTT	616
Dd	404	AYMTNGRTEQYWKNTATVLDSDGSFYMFYSKL.RVQKSTWERGSLFACSVHIEGIHNHLT	463
Oy	617	QKSLSLSPG 625	
Dd	464	TKTISRSLG 472	
 RESULT 28 099LJ31			
ID	099LJ31	PRELIMINARY; PRT; 468 AA.	
AC	099LJ31;		
DC	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, last annotation update)		
Dc	Similar to RIKEN CDNA 181060C009 gene.		
Dc	Mus musculus (Mouse).		
Oc	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Ox	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
Ox	NCBI_Taxid=10090;		
Rn	[1]		
Rf	SEQUENCE FROM N.A.		
Ra	Strausberg R.;		
Rl	Submitted (FEB.-2001) to the EMBL/GenBank/DBJ databases.		
Rl	EMBL; BC003878; AA03878.1; --		
Dd	PDB; 2AP2; 24-NOV.-99.		
Dd	InterPro; IPR007110; Ig-like.		
Dd	InterPro; IPR003006; Ig_MHC.		
Dd	InterPro; IPR003596; Ig_V.		
Dd	Pfam; PF00047; Ig_3.		
Dd	SMART; SM00406; IG; 1.		
Dd	SMART; PS50835; IG LIKE; 4.		
Dd	PROSITE; PS00290; IG_MHC; 1.		
Dd	PROSITE; PS00290; IG_MHC; 1.		
Sq	SEQUENCE 468 AA; 51661 MM; 96352328B332ADB CRC64;		
 Query Match 24.4%; Score 832; DB 11; Length 468; Best Local Similarity 35.5%; Pred. No. 6.8e-55; Matches 217; Conservative 65; Mismatches 140; Indels 190; Gaps 211			
Oy	30	LKKKDPTELTCAS--OKSKIQPHWNKNNOIKTLGNQ-----SFLTKGPSKLN 77	
Dd	30	LVPFGASVXLKSLCSAFNFKOSLMHW-----VKORPEKGLEMIWGIDPEDGETKYAPFQ 84	
Oy	78	DRDARSRLMDGNGPFLIIKN.KTLEDSPYIC.EVBDQKEBVQLLVPLGTANSPTHLIQG 137	
Dd	85	DKAATTATDTSNWTA-YLQSLSTSSEDIAIYYC-----ARNLTYG-----GYDYWGQ 130	
Oy	138	SLTLTLSPGSSPSVQCSPKGRKNI.QGKTLVSVOLELQDSPGTWTCTVLQNOKVFEKI 197	



```
Db 131 GTTITVSSAKTTAPSVYELAR-----VCGDTT----- 157
Qy 198 DIVVLAFOKASSIVYKKEGOVEFSEPLAFVTEKLTGSGELMWOERASSKSMITFDLK 257
Db 158 -----GSSVTL--GCLVKGTFP-----EPVT-----LTWNGSLSSG----- 187
Qy 258 NKEVSVKRVTDOPKQLQMGKKLPLHLTLPLQALPQ--YAGSGNLTALAEKTKLHQEVNLV 315
Db 188 -----VH-TFPRAVLQSDLYTLSSSVT----- 207
Qy 316 VMRAQLOLOKULTCEVWGPTSPKLMSTLENKEAVNSKREKRPVWNLNPEAGMOCCLSDS 375
Db 208 VTSTWPSQSIITCVNVAHPAS-----STVDDKIEP-----R 238
Qy 376 GOVLLESNIKVLPTWSTVPEPKSCDKTHTCPP--CPABELLGSPVFLFPKPKDTLMIS 433
Db 239 G-----PT-----IKP-----CPCKCPAPNLLGSPSVFPPKIKIDVLMIS 275
Qy 434 RTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKRPREOYNSTIRRVSVLTVLDHOML 493
Db 276 LSPMTTCVVVDVSEDDPVOGISMFWNNVEVLTAAQTQTHREDYNSLTRVVSALPIQHODMM 335
Qy 494 NGKEYKCKVSNKALPAPIEKITISKAKGQPREPOVYITLPPSDDELTKNQVSLTCLVKGYFP 553
Db 336 SGKEKCKCVNNKALPAPIERTISKPKGSVRAQYVVLPPPEEMTKKQVTLTCVWTDPMF 395
Qy 554 SDIAVEMESNGOPENNYKTPTPVLDDSGSPFLYSKLTVDKSRMOQGNVFCSVNHEALHN 613
Db 396 EDIYEMTNNGKTELNTKTEPVLDDSGSYFMYSKLRVEKKWVERNSYSCSVNHEGLHN 455
Qy 614 HYTKSKLSLSPG 625
Db 456 HHTTKSFSTRTPG 467
```

## RESULT 29

```
Q09125 PRELIMINARY; PRT; 473 AA.
ID 09125;
AC 09125;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to RIKEN CDNA 181006009 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003888; AA03888.1; -
DR HSSP; P01842; 7EAB.
DR InterPro; IPR007110; Ig_1like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00406; Igv.1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 473 AA; 52449 MW; BE9889B7986DA155 CRC64;
```

Query Match 23.7%; Score 810.5; DB 11; Length 473;

Best Local Similarity 47.4%; Pred. No. 3e-53; Indels 59; Gaps 7;

Matches 165; Conservative 47; Mismatches 77;

```
Qy 285 PQALPOYAGSNTLALAEKTKLHQEV-----NLVNRATQLOKULTCEVWGPTSPKLM 339
Db 177 PEPVTLTNNGSISGVTTPRAVLQSDLYTLSSSVTVSSTPSPGSIICVNAHPAS----- 232
Qy 340 LSLKLENKAYSKREKVPVWNLNPEAGMOCCLSDSGVLLLESNIKVLPTWSTVPEPKSC 399
Db 233 -----STVDDKIEP-----RG-----PT-----IKP----- 249
```

```
Qy 400 DKHTTCTPP--CPABELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWY 457
Db 250 -----CPCKCPAPNLLGSPSVFPPKIKIDVLMISLSPMTTCVVVDSEDDPVOGISM 304
Qy 458 VDGVEVHNAKTKRPREOYNSTIRRVSVLTVLDHOMLNGKEYKCKVSNKALPAPIEKITISK 517
Db 305 VNNVEVLTAAQTQTHREDYNSLTRVVSALPIQHODMWSGKEFKCKVNNKALPAPIERTISK 364
Qy 518 AKGQPREPOVYITLPPSDDELTKNQVSLTCLVKGYFSDIVEMESNGOPENNYKTPPVL 577
Db 365 PKGSVRAQYVVLPPPEEMTKKQVTLTCVWTDPMFEDIYEMTNNGKTELNTKTEPVL 424
Qy 578 DSDGSPFLYSKLTVDKSRMOQGNVFCSVNHEALHNHYTKSKLSLSPG 625
Db 425 DSDGSYFMYSKLRVEKKWVERNSYSCSVNHEGLHNHHTTKSFSTRTPG 472
```

## RESULT 30

```
Q091205 PRELIMINARY; PRT; 473 AA.
ID 091205;
AC 091205;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
GN AU044919.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010327; AA010327.1; -
DR MGD; MGI:2144967; AU044919.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000345; Cytc_heme_BS.
DR InterPro; IPR007110; Ig_1like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00406; Igv.1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
DR Hypothetical protein.
SQ SEQUENCE 473 AA; 51946 MW; CF625F008932AF12 CRC64;
```

Query Match 23.4%; Score 800.5; DB 11; Length 473;

Best Local Similarity 33.5%; Pred. No. 1.7e-52; Indels 179; Gaps 15;

Matches 204; Conservative 69; Mismatches 157;

```
Qy 30 LGKKGDTVELTCTAS--QKSKIQFHWKNSNQIKILG-----NQGSP-----LTGSPSKL 76
Db 30 LVKPGGSRKLSGASGTFSDYGMHWRAQAPBKGLAVAVINSSTTIYADTVKGRFTI 89
Qy 77 NDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEBVOALLVGLTRANSDFHLLQG 136
Db 90 S-RDNARKNTLP-----LQMTSLRSEDTAMYYCARBELMLRID-----YWG 128
Qy 137 QSLTLTLESPPGSSPSVQCRSPKGNIQGGKTLVSQLELDQDSGTWCTVQLQNKVYEFK 196
Db 129 QGTTITVSSAKTTPPSVPLAPGCGDTTG----- 157
Qy 197 IDIVLAFOKASSIVYKKEGOVEFSEPLAFVTEKLTGSGELMWOERASSKSMITFDL 256
Db 158 -----SSVTL--GCLVKGTFPESVTVI-----WNGSLSS----- 186
Qy 257 KNKEVSVKRVTDOPKQLQMGKKLPLHLTLPLQALPOYAGSNTLALAEKTKLHQEVNLV 316
```



```
Db 187 -----VH-TFPALL-----QSGLYTMSSSVTV 207
Qy 317 MEATOLQKLTCEVWGPRTSPKMLSLKLENKAKVSKREKPVWVNLPEAGMQLLSDSG 376
Db 208 PSTSTPSCVITCSVAHPAS-STTVDKLE-----PSCPISTINP----- 245
Qy 377 QVLLSNNIKVLPWSTPVEPKSCDKTHTCPCPAPPELLGSPVFLFPKPKDTLMISRP 436
Db 246 -----CPCKECHKCPAPLLEGSPVFIFFPPIKQVLMISLTP 283
Qy 437 EUTCYVVVVDVSHDEPEKFWVYDGVENNAKTPREBOYNSTRYVSVLTJLHOMLNGK 496
Db 284 KYTCVVVVDSEDDPDQVISMFWNNVEVHTAQTOTREDYNSTIRVVSALPIQHDWMSGK 343
Qy 497 EKCKVSNKALPAPIEKTSKAKGPREPOVYTLPPSRDELTKNOVSLTCLVKGFPYPSDI 556
Db 344 EKCKKNNKDLPSPIERTISKIKGLVAPQVYILPPAPQLSRKQVSLCLVAGNRPDI 403
Qy 557 AVEWESNGOPENNYKTPPVLDSDGSPFLYSKLTVDKSRMQOQNVFSCVMHEALHNHYT 616
Db 404 SVEWTSNGHTENYKDTAPVLDSDGYFIYSKLDIKTSKMEKTDSPSCVNRHGLKNYTL 463
Qy 617 QKSLSLSPG 625
Db 464 KKTISRSPG 472

RESULT 31
Qy 09R3H6 PRELIMINARY; PRT; 474 AA.
ID 09R3H6
AC 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
GN A0044919.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 1)
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC025447; AAH25447.1; -.
DR MGD; MG:2144367; A0044919.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000345; Cytochrome_BS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS00835; IG LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KM Hypothetical protein.
SQ SEQUENCE 474 AA; 51748 MW; 8608B576CD2874A CRC64;

Query Match 23.4%; Score 799.5; DB 11; Length 474;
Best Local Similarity 33.2%; Pred. No. 2,1e-52;
Matches 209; Conservative 71; Mismatches 155; Indels 195; Gaps 18;
```

```
Db 121 -----YG-----DYDDWKQGAIVTVSSAKTTPSPSYPLAPGCGDPTG----- 158
Qy 176 LODSGTWTCTVLQONKKEFKFIDIVLAFQCAASSIVYKKEBOQEFSPPLAFYHEKLTGS 235
Db 159 -----SSVTL-----GCLVYGFPESVTVT----- 178
Qy 236 GELMWQAERASSSKSWITFDLKNKEVSVKVTVQDPKLGKKLPLHLTLPOALPYAGSG 295
Db 179 -----MNSGLSSS-----VH-TFPALL----- 195
Qy 236 NLTLAEKTKLHDEVLVVMRATOLQKLTCEVWGPRTSPKMLSLKLENKAKVSKRE 355
Db 196 -----QSGLYTMSSSVTVPSSSTWPEQYTCVAHPAS-STTVDKLE-----PS 238
Qy 356 KPVWVNLPEAGMQLLSDSGQVLLSNNIKVLPWSTPVEPKSCDKTHTCPCPAPPELLG 415
Db 239 GPISTINP-----CPCKECHKCPAPNLEG 263
Qy 416 GPSVFLFPKPKDTLMISRTPEVTCVVVDVSHDEPEKFWVYDGVENNAKTPREBOY 475
Db 264 GPSVFIFFPPIKQVLMISLTPKAVTCVVVDSEDDPDQVISMFWNNVEVHTAQTOTREDY 323
Qy 476 NSTYRVVSVLTJLHODWLNKKEKCVSNKALPAPIEKTSKAKGPREPOVYTLPPSRD 535
Db 324 NSTIRVVSALPIQHDWMSGKEPKCKVNNKDLPSPIERTISKIKGLVAPQVYILPPAP 383
Qy 536 ELTKQVSLTCLVKGFPYSDIAVEWESNGOPENNYKTPPVLDSDGSPFLYSKLTVDKSR 595
Db 384 QLSKDVSLTCLVVGFGNDISVWTSNGHTEENYKDTAPVLDSDGYFIYSKLDIKTSK 443
Qy 596 MOQGVNPSQVMHEALHNHYTOKSLSLSPG 625
Db 444 WEKTDSPSCVNRHGLKNYTLKTIISRSPG 473
```

```
RESULT 32
Qy 096EY0 PRELIMINARY; PRT; 613 AA.
ID 096EY0
AC 096EY0;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN 1)
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC011857; AAH11857.1; -.
DR PIR; S15590; S15590.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KM Hypothetical protein.
SQ SEQUENCE 613 AA; 67273 MW; 31214203FB8421E7 CRC64;

Query Match 12.2%; Score 417.5; DB 4; Length 613;
Best Local Similarity 22.2%; Pred. No. 3.7e-23;
Matches 151; Conservative 117; Mismatches 280; Indels 131; Gaps 24;
```



```

OY 62 LGND--GSFLTKPEPKLMDRADSRSL--WDQGNFPLIIKLIKIEDSDTYICEVEDQKE 116
DB 62 KGLEWIGRIYTSSTNNYPSLKSRYTMSVDTSKNQFSLKLSVAADTAAYVC---ASQ 117
OY 117 EVOLLVFLTLANSDBHLQGSGLTITLESPPGSSPS---VQCSSPRKNIQGGKTLSSVS 172
DB 118 PWEIPLTVL-----FYWGGQTLTVTSSGASASATLTPPLVSC----- 153
OY 173 QLEIADSGTWTCTVLQNOKKEKVIDIVLAFQ--KASSIVYKKEGQEVFSFPLAFTVEK 231
DB 154 ----ENSPSDTSSVAAGLADFLPDSITTFPMKTKNNNDISTG-----FPSVLKGGK 203
OY 232 LTGSGELMWQABRASSSKSWITFDLKNKESVSKRVTDPKLQMGKPLHL--TLPOALP 289
DB 204 YAATSQVLLPESKDV-----MQGDEHVCKVQHPNKNKEKNVLPVIAELPKVVS 253
OY 290 QYA-----GSGNITLALBAKTKLHQBENLVVNR-----ATQQRKLTCEWNGPT 334
DB 254 VFVPRDGFENPNPKSKLICOATGFSPPQIQVSWLREGKQVSGVTTDQVGAKEKSGPT 313
OY 335 SPKMLSLKLENKEKAVSKREKPVVVLNPEAGMQCLSDSGQVLE--SNIKVLPTWSTP 393
DB 314 TYKNTSTLTITKESD-----WL---SOSMFTCRVDHRLGTFQOANASMCVPDDQTA 360
OY 394 VEPKSCDKHTTCCPCAPABELLGGPSVFLFPKPKDPTLMIISRTPEVTCVVDVSHEDPEVK 453
DB 361 IR-----VFALPPS--FASIFPLTKSTKLCTCLVDTLTYD--SVT 395
OY 454 FNNVVDGVEVNNATKPREBOYNSYRVVSVLTVLHQDLNKGKEYKCVSKALPAPLEK 513
DB 396 ISWRONGEAVKTHNTNISESHPNATFSAVGASLCEEDMNSGERFTCTVHTDLPSPLKQ 455
OY 514 TISKAKGP--REPOVYTLPPSRDEL--KNQVSLTCLVKGFPSPDIIVAWESNGQP--ENN 569
DB 456 TISPKGVALRPPDVLVLPARBOULNRESATITCLTVGFSPADVPVQMGKGPLSEK 515
OY 570 YKTPPVLD--SDGSFFLYSKLTVDKSRWQGNFSCSVMEHALHNHYTKSLSPGLQ 627
DB 516 YVTSAPMPEPQAPGRYFAHSILITVSEBEMNGETVTVVAHEALPNRTERTVDSK---- 571
OY 628 LDFTCAEAGDGLDGLWTT 646
DB 572 -TEGEVSADEEGFENLWAT 589

RESULT 33
O7TMT6 PRELIMINARY; PRT; 614 AA.
AC O7TMT6;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;
RX MEDLINE=22386257; PubMed=12477932;
RA Strauberg R., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Muljaby S.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

```

```

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez J.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzyzanski M.I., Skalka U., Smalhus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;
RA Strauberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053409; AAH53409.1; -.
KW Hypothetical protein.
SQ SEQUENCE 614 AA; 67746 MW; 839BAF3B8D124F89 CRC64;

Query Match 12.24; Score 416; DB 11; Length 614;
Best Local Similarity 22.74; Pred. No. 4.8e-23;
Matches 156; Conservative 101; Mismatches 233; Indels 198; Gaps 27;

OY 30 LGKKGDTVELTCTASQKSKIQFHMKX-----SNOIKILGN-----QGSFLTKGP 73
DB 30 LVKPGASVKISKAS--GYAFSSSMWVVKORPKGLWIRVYPGDGTNNGKF--KGG 86
OY 74 SKLN--DRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEVQLVFGTLANS DTH 132
DB 87 ATLADKSSSTAWQ-----LSSLTSDSAVYFC--ANDYSSYFAVWG----- 129
OY 133 LLOGQSITLTLESPP--PGSSPSVQCRSP--RGKNIQGGKTLSSVLEIADSGTCTVLQNL 189
DB 130 --QGLTVTVAESQSFPNVPFLVSCSPSLSDKLVLMGCLARDFLSTISFTNN----- 181
OY 190 QKXVEPRIDIVLAFQKASSIVYKKEGQEVFSFPLAFTVEKLTGSGELMWQABRASSSK 249
DB 182 -----YONNAEVIQ-----GIR 193
OY 250 SWITFDLKNKESVSKRVTDPKLQ-----GKKPLHLTLPOALPOVAGSG 295
DB 194 TPTPLRTGKRYLNTQVLSPKSILBSGDEYLVCKIHYGKNNDLHVP----- 243
OY 296 NLTLALBAKTKLHQBENLVV-----WRATQLOKNTLCEWNGPSPKMLSLKENKEA 349
DB 244 -----AVALNNPVNVVFPVPRDGFSGPAPRKSXKLICEATVFTPRKPTVSWLKDGKLV 295
OY 350 KVSREKPVVNL-----PEA-----GMNQCLSDSGQVLESNIKVL 388
DB 296 ESGFTDPEVTIENKSGTPTQYKVIISTLTISEIDMLNLTVCVDRHGLTFK----- 348
OY 389 TWSTPVEPKSCDKHTTCCPCAPABELLGGPSVFLFPKPKDPTLMIISRTPEVTCVVDVSH 448
DB 349 -----ANSTCAASPSDIL-----TFTIPSPAD--IFLSASANTCLVSNLATY 392
OY 449 DPEVKENVYDGVENHNAKTPREBOYNSYRVVSVLTVLHQDLNKGKEYKCVSKALP 508
DB 393 E--TLNISWASQSEPLETKIKIMESHPNNGTFPSAKGASVCEVDMNRKKEVCTVTRDLP 451
OY 509 APIEKTISKAKGPPE-----POVYTLPPSRDEL--KNQVSLTCLVKGFPSPDIIVAWES 562
DB 452 SPKKKFTSK---PNEVHKHPPAVVLLPAPREOLNRESATVTVCLVKGFPADIIVQWMLQ 507
OY 563 NQO--PENNYKTPPVLD--SDGSFFLYSKLTVDKSRWQGNFSCSVMEHALHNHYTK 618
DB 508 RGQLLQGEKVTVAAPPEPAPRPFYTHSLITVTEBMSNGEFTVTVSHALPHLVTER 567
OY 619 SLSLSPGLQIDFTCAEAGDGLDGLWTT 646
DB 568 TVDKS-----TEGEVVAABEGFENLWAT 590

RESULT 34
O8WUK1 PRELIMINARY; PRT; 613 AA.

```



AC Q8WUK1;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RC Struhsberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RL EMBL, BC020240; AAH20240.1; -  
 DR PIR, P10120; P10120.  
 DR PIR, S15590; S15590.  
 DR InterPro, IPR007110; Ig-like.  
 DR InterPro, IPR003006; Ig\_MHC.  
 DR InterPro, IPR003596; Ig\_V.  
 DR Pfam, PF00047; Ig\_5.  
 DR SMART, SM00406; IGV, 1.  
 DR PROSITE, PSS0835; IG\_LIKE, 5.  
 DR PROSITE, PS00290; IG\_MHC, 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 613 AA; 67296 MW; 60C7F5950671E315 CRC64;

Query Match 11.9%; Score 407.5; DB 4; Length 613;  
 Best Local Similarity 22.6%; Pred. No. 2,1e-22;  
 Matches 150; Conservative 119; Mismatches 252; Indels 143; Gaps 27;

25 GNKVVIGKKGDTVELTCTAS--QKSIQFHW-----KNSNOKILIGNGS-----FLTK 71  
 27 GGGVV--QGRSRIRLSCAASGFFSSYGHWRAQBGKLEWVAVSYDGSNKYVADSK 84  
 72 GSKLNDRAADSRSLWDQGNFLLIKLKIEDSDTYICEVDQKEVQLVFGLTANSPDT 131  
 85 GPFITIS-RNRSKNTLYLQWN-----SLRAEDTAVYYC-AKDMSEGV-----ET 125  
 132 HLQOGSLTLTLESPPGSSPS---VQCRSPRGNKIQGKTLVSQLELDGSGTWCTYL 187  
 126 FDIWGGTWTVSSGSASAPLTFPLVSC-----ENSPSDTSVA 164  
 188 QNOKKVEFKIDIVLAFO--KASSIVKKEGEVESEFLAFTYEKLTSGLMWQAERAS 246  
 165 VQGLADFLPDSITTSWTKRNSDISSTG-----FESVLKAGKTAATSOVLPSKDV- 217  
 247 SSKSWITFDLKNNKESVKEVTQDPKLQMGKULPLHL--TLPOALPOYA-----GSGNL 297  
 218 -----MGGDEHNVCKVQHPNCKEKNVPLPIALPLPKVSVFVPRDGFQNPXK 268  
 298 TLALNAKTKLHDEVNVLVNR-----ATOLQKNLTCEVWGFPSPKMLSLKLENKEA 349  
 269 SKLICQATFSPROIQVSWLREGQVGSVTTDQVQAEAKESGPTTYKVTSLTIKESD- 327  
 350 KVSKEKPVWVLPNPEAGMQLLSDSGOVLLE--SNIKVLPWSTPVEPSCOKHTCPFC 408  
 328 -----WL---SOSMFTCRDHRGLTFQOVMASMCVPPDDTAR----- 362  
 409 PAPELLGSPSVLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFMVVDGEVYNAKT 468  
 363 -----VFAIPPS--PASIFLTKSTLTCLVLDLTYYD--SVTISWTRNGEAVKHTT 410  
 469 KREGEQNSTYRVSVLVTLHODMLNGEKYKCKVSNKALPADIETISAKQOP--REPOV 527  
 411 NISSEHPNATFSVAGSASICEEDMNSGERFTCTVHTDLPSPDKQITSPKGVALLHRPDV 470  
 528 YTLPEPRDLT--KNQVSLTCLVKGFPSPDIAYEMENQOP--ENNYKTPPLVD--SDS 582  
 471 YLLPFAREQDNLNESAFTITCLVTFGSPADVPVQMMQRGQPLSPKTVISAPPEQAPGR 530  
 583 FFLYSKLTVDKSRWQGNVPSGVNMEALHNHYTQKSLSPGLQIDETCAEADQDELGD 642

DB 531 YFAHSILTVSEENMTGERTYCVVAHEALPNRTERTVDS-----TEGEVSADSEGFEN 585  
 QY 643 LMTT 646  
 DB 586 LMAT 589

RESULT 35

ID 077596 PRELIMINARY; PRT; 86 AA.  
 AC 077597;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).  
 OS Mandrillus sphinx (Mandrill) (Papio sphinx).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Mandrillus.  
 NX NCBI\_Taxid=9561;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98320644; PubMed=9656488;  
 RA Harris E.B., Disotell T.R.;  
 RT "Nuclear gene trees and the phylogenetic relationships of the  
 RT mangabeys (primates: Papionini)."  
 RL Mol. Biol. Evol. 15:892-900(1998).  
 CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL  
 CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY  
 CC SIMILARITY).  
 CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
 CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.  
 DR EMBL, AF057386; AAC25130.1; -  
 DR HSP, P01730; ICDY.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR InterPro, IPR007110; Ig-like.  
 KW Immunoglobulin domain; T-cell; MHC; Transmembrane.  
 FT NON\_TER 1  
 FT DOMAIN 1  
 FT DOMAIN 20  
 FT DISULFID 49 78  
 FT NON\_TER 86  
 FT NON\_TER 86  
 SQ SEQUENCE 86 AA; 9406 MW; 2BD97A95E19582AB CRC64;

Query Match 11.7%; Score 400; DB 6; Length 86;  
 Best Local Similarity 89.4%; Pred. No. 4.8e-23;  
 Matches 76; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 107 YICEVEDQKEEVOLVFLGTANSDTHLLQGGSLTLTLESSPGSSPSVQCRSPRGNKIQG 166  
 DB 1 YICEVEDQKEEVOLVFLGTANSDTHLLQGGSLTLTLESSPGSSPSVQCRSPRGNKIQG 60  
 QY 167 KTLVSQLELDGSGTWCTVYQONK 191  
 DB 61 RTLVSQLELDGSGTWCTVYQONK 85  
 RESULT 36  
 ID 077597 PRELIMINARY; PRT; 86 AA.  
 AC 077597;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).  
 OS Mandrillus leucophaeus (Drill) (Papio leucophaeus).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Mandrillus.



```

OX NCBI_TaxID=95568;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RL mangabeys (primates: Papionini).";
RM Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057387; AAC25131.1; -.
DR HSSP: P01730; 1CDY.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR InterPro: IPR007110; I9-like.
DR Immunoglobulin domain; T-cell; MHC; Transmembrane.
KW
FT DOMAIN 1 19 IG-LIKE V-TYPE DOMAIN.
FT NON TER 1
FT DISULFID 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON TER 86
SQ SEQUENCE 86 AA; 9406 MW; 2BD97A9EEL19582AB CRC64;

Query Match 11.6%; Score 400; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 4.8e-23;
Matches 76; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 107 YICEVEDQKEEVQLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCRSPRGKNIQGG 166
DB 1 YICEVEDQKEEVQLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCRSPRGKNIQGG 60
OY 167 KTLVSQLELDSDGTWCTVLOXOK 191
DB 61 RTLSPOLERQDSGTWCTVVSQDOK 85

RESULT 37
OY 077594 PRELIMINARY; PRT; 86 AA.
ID 077594;
AC 077594;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE (Fragment).
OS Cercopithecus mitis.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=36225;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RL mangabeys (primates: Papionini).";
RM Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057381; AAC25125.1; -.
DR HSSP: P01730; 1CDY.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR InterPro: IPR007110; I9-like.
DR Immunoglobulin domain; T-cell; MHC; Transmembrane.
KW

```

```

FT NON TER 1 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DISULFID 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON TER 86
SQ SEQUENCE 86 AA; 9417 MW; 371CA39EF58182AB CRC64;

Query Match 11.6%; Score 397; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 8.2e-23;
Matches 76; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 107 YICEVEDQKEEVQLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCRSPRGKNIQGG 166
DB 1 YICEVEDQKEEVQLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCRSPRGKNIQGG 60
OY 167 KTLVSQLELDSDGTWCTVLOXOK 191
DB 61 RTLSPOLERQDSGTWCTVVSQDOK 85

RESULT 38
OY 077599 PRELIMINARY; PRT; 86 AA.
ID 077599;
AC 077599;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE (Fragment).
OS Theropithecus gelada (Gelada baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopithecinae; Theropithecus.
OX NCBI_TaxID=9565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RL mangabeys (primates: Papionini).";
RM Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057389; AAC25133.1; -.
DR HSSP: P01730; 1CDY.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR InterPro: IPR007110; I9-like.
KW Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT NON TER 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DISULFID 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON TER 86
SQ SEQUENCE 86 AA; 9420 MW; 2BD97A98574EE9AB CRC64;

Query Match 11.6%; Score 397; DB 6; Length 86;
Best Local Similarity 88.2%; Pred. No. 8.2e-23;
Matches 75; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

OY 107 YICEVEDQKEEVQLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCRSPRGKNIQGG 166
DB 1 YICEVEDQKEEVQLVFGLTANSPTHLQGSLLTLESPPGSSPSVQCRSPRGKNIQGG 60
OY 167 KTLVSQLELDSDGTWCTVLOXOK 191
DB 61 RTLSPOLERQDSGTWCTVVSQDOK 85

```



RESULT 39  
 O8VCX7 PRELIMINARY; PRT; 613 AA.  
 AC O8VCX7;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN IGH-6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=salivary gland;  
 RA Strubberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC018315; AAH18315.1; -  
 DR MGD; MGI:96448; Igh-6.  
 DR InterPro; IPR007110; Ig\_1like.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003596; Ig\_V.  
 DR Pfam; PF00047; IgV\_1.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PSS00835; IG\_LIKE; 5.  
 DR PROSITE; PSS00290; IG\_MHC; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 613 AA; 67855 MW; 41A9384DD4C22862 CRC64;

Query Match 11.6%; Score 396.5; DB 11; Length 613;  
 Best Local Similarity 22.7%; Pred. No. 1.5e-21;  
 Matches 157; Conservative 100; Mismatches 227; Indels 209; Gaps 27;

30 LKKKGDVTELTCTAS-----QKSAIQFMKNSNOIKILGNQS--FLTKGPS 74  
 30 LMKPASVXISKATGYTFSSYWIEMVKORPGHGLEWIG---ELIPGSGTNYNEKFKG 85  
 75 KLNDAAD-SRRLMOGNFPLIIKLIKIDSDTYICEVDQKEVQL-----VFGTLA 127  
 86 KATFTADTSSNTAYNQ-----LSLSTSDSAVYVC-----ARRLGWYFDVWG--- 128  
 128 NSDTHLQOGSLTLTLESP--PGSSPSVOCRRP-RGKNIQGGKTLVSQLELDGSGWTTC 184  
 129 -----AGTYTVSSSSQSPFNVPFLVSCSPSLDKULVAMGCLARDLPBETIFTNN- 180  
 185 TVLQNKQKVEFKIDIVLAFQKASSIVYKKEGOVEFSPFLAFTVEKLTGSGELWQAEK 244  
 181 -----YQNTVEVIG----- 189  
 245 ASSSKSWITFDLAKKEVSKRYTQDPKLOM-----GKKLPHLTLTPOALPO 290  
 190 -GITFTPLRTGKGLATSOVLSPKSLIEGSDVYLCKIYGGKRNRLHVPPI----- 242  
 291 YAGSGNLTALBAKTKLHQEVNLVY-----MRATQLOKNTLCEVWGPSTPKMLSLKL 344  
 243 -----AAEKNPNVNVFPPPDGSGAPRPSKILICETATNTPPRTIVSMUK 289  
 345 ENKEAKVSKREKRPVWVLN---PEA-----GMQOCLSDSGOVLLESN 383  
 290 DQGLVSGFTTDPVLTENKSGTPQYKIVLSTLISIDWLNMYTCRDHNGHGLFLK-- 347  
 384 IKVLTWTSFVPEPKSCDKHTKTPPCAPAPELLGGPSVFLFPPPKKOTLMSRTPEVTCVV 443  
 348 -----NWSSTCAASPSDIL---TFITPSPFAD-IFLSKSNLTCIVS 386  
 444 DVSHDEPEYKFMVYDGVVEVNAKTRPREQYNSYTVRVSVLTVLHODMLNKEYKCVS 503  
 387 NLATYE-TLNIWASQSGEPLETKIKIMESHNGFTSAGKVASVCEVDNNKKEVCTVT 445  
 504 NKALPAPIEKTISKAGOPRE-----POVYTLPPSRDELTL-KNOVSLTCLVKGFPSPDIA 557  
 446 HRDLSPQPKFKISK-----PNEVHKIPRAVYLLPPRAREQNLRESATVTCIVLGVSPADIS 501

OY VESWESNGO--PENNYKTPVPLD--SDGSFPLYSKLTVDKSRMOQGNVPSGVMEALHN 613  
 DB VQMLORGQULPOEKTYTSAPMPGPAPGFYFTHSLITVTEERMSGSEYTTCVGHGALPH 561  
 OY 614 HYTKSLSPQLQDETCAEAKODELDGLMTT 646  
 DB 562 LVTERTVKSS-----TEGEVNAEEBGFENIMWT 589

RESULT 40  
 ID 077595 PRELIMINARY; PRT; 86 AA.  
 AC 077595;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).  
 OS Cercopithecus galieri chrysogaster.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Cercopithecidae.  
 NC NCBI\_TaxID=75569;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98120644; PubMed=9656488;  
 RA Harris E.E., Disotell T.R.;  
 RT "Nuclear gene trees and the phylogenetic relationships of the  
 mangabeys (primates: Papionini).";  
 RL Mol. Biol. Evol. 15:892-900(1998).  
 CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL  
 RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY  
 SIMILARITY).  
 CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE IMMUGLOBULIN SUPERFAMILY. CONTAINS  
 ONE V-LIKE AND 1 C2-LIKE DOMAIN.  
 DR EMBL; AF057382; AAC25126.1; -  
 DR HSSP; P01730; ICXY.  
 DR GO; GO:0016021; C:Integral to membrane; IEA.  
 DR InterPro; IPR007110; Ig\_1like.  
 KW Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.  
 FT DOMAIN 1  
 FT <1 19 IG-LIKE V-TYPE DOMAIN.  
 FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.  
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT DISULFID 49 78 BY SIMILARITY.  
 FT NON TER 86  
 SQ SEQUENCE 86 AA; 9419 MW; A9D97A9E19582BE CRC64;

Query Match 11.6%; Score 395; DB 6; Length 86;  
 Best Local Similarity 88.2%; Pred. No. 1.2e-22;  
 Matches 75; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 107 YICEVEDQKEVQLVFGLTANSPTHLLQGSLLTLESPPGSSPSVOCRRPRGKNIOGG 166  
 DB 1 YICEVEDQKEVELLVFGLTANSPTHLLQGSLLTLESPPGSSPSVOCRRPRGKNIOGG 60  
 OY 167 KTLVSQLELDGSGTWTCTVLANOK 191  
 DB 61 RTLVSQLELDGSGTWTCTVLANOK 85

RESULT 41  
 ID 096BB9 PRELIMINARY; PRT; 597 AA.  
 AC 096BB9;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).



OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=B-cell;  
 RA Strauberg R.;  
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC015760; AAH15760.1; -  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_TV.  
 DR Pfam; PF00047; Ig; 5.  
 DR SMART; SM00406; IG; 1.  
 DR PROSITE; PS00835; IG LIKE; 5.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 597 AA; 65039 MW; 4FC3AD8ECE263D9 CRC64;

Query Match 11.5%; Score 392.5; DB 4; Length 597;  
 Best Local Similarity 22.4%; Pred. No. 2.9e-21;  
 Matches 143; Conservative 116; Mismatches 248; Indels 131; Gaps 24;

30 LGKGGDTVELTCTAS--QKSIQPHKNSNQK-----ILNGSGFL---TKGPKL 76  
 30 LVOPGSGIRLSCASGFSFSSYAMWVRQAPGKLEWVAISGSGSTYVADSVYGRFTI 89  
 77 NDRADSRSLMDQGFLLIKLKIEDSDTYICEVEDQKEVQLVPLGLTNSDTHLLOG 136  
 90 S-RDNRSLTLYQNN-----SLRADTAIVYCAKDPNG-----YASGVYREDYWG 135  
 137 QSLTLTLSPGSSPS---VQCRSPRKNIGGKTLVSQLELDGSGTWTCTVLOQNK 192  
 136 QGLTVTVSSGSAATLPLVSC-----ENSPSDTSSVAVGCLA 174  
 193 VEFKTDIVLAFQ-KASSIVYKKEGQVEFSPPLAFYTEKLTGSGELMWQERASSSKSW 251  
 175 QDFLPDSITFSWKYKNSDISSTRG-----FPSVLRGKYAAVSQVLLPSKDV----- 222  
 252 ITFDLKNKEVSKRYTODPKLQMGKLLPLHL--TLQALPOVA-----GSGNLTAL 302  
 223 ---WQGDENHYVCVQHPNGKKNVPLVYAEIPRYVSFVPRDPFGAPRKSKLIC 278  
 303 AKTGKHOENVLVVR-----ATOLQNLTCVWGPTSPKMLSLKENKEAKVSKR 354  
 279 QATGSPRQIQVSMRBEKQVGSVTTDQVQAEKSGPTTYKATSTLTIKESD----- 332  
 355 EKPWVVLNPEAGMOCCLISDSQVLE-SNIVKLPWTASTPVEPKSCDKTHTCPPCAP 413  
 333 ---WL---SOSMPTCRVDHRLGTFQONASSMCVDDQDTAIR----- 367  
 414 LGGSEVFLPPPKOTLMISRTPEVTCVAVDVSHDEPKFMWYDVGVVNAKTKPRE 473  
 368 ---VFAPIDPS-FASIFLTSTKLTCLVTDLTVD-SVTSWTRONGEAAYKHTNISES 420  
 474 QYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIEKTISKAGOP-REPOVYTLPP 532  
 421 HPNATFSAVGEASICEBDMNGSERFTCTVHTDLPSPKQITSRKQVLAHRPDVYLLP 480  
 533 SRDELTKVQVSLTVKGFYPSDIAYMESNGOP--ENNYKTTPEVLD--SDGSFFLYS 587  
 481 AREQNLNLESATITCLVTFGSPADVFVQMQRGQPLSPKQVTSAPMPEQAPGRYFAHS 540  
 588 KLTVDKSRMGOQGNVSGVMHEALHNHTOKSLISLSPG 625  
 541 ILTVSEEMTGETYTCVVAHEALPNRYTERKVDKSTG 578

RESULT 42  
 077598 PRELIMINARY; PRT; 86 AA.  
 AC 077598;  
 DT 01-NOV-1998 (TReMBLrel. 08, Created)

DT 01-NOV-1998 (TReMBLrel. 08, last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, last annotation update)  
 DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)  
 DE (Fragment).  
 OS Papio sp. (baboon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
 OC Cercopitheciinae; Papio.  
 OC NCBI\_TaxID=61183;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98320644; PubMed=9656488;  
 RX Harris E.E., Disotell T.R.;  
 RT "Nuclear gene trees and the phylogenetic relationships of the  
 mangabeys (primates: Papionini).";  
 RL Mol. Biol. Evol. 15:892-900(1998).  
 CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL  
 RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY  
 SIMILARITY).  
 CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
 ONE V-LIKE AND 1 C2-LIKE DOMAIN.  
 DR EMBL; AF057388; AAC25132.1; -.  
 DR HSSP; P01730; ICDY.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR InterPro; IPR007110; IG-like.  
 KW Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.  
 FT NON TER 1  
 FT DOMAIN 1  
 FT <1 19 IG-LIKE V-TYPE DOMAIN.  
 FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.  
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT DISULFID 49 78 BY SIMILARITY.  
 FT NON TER 86  
 SQ SEQUENCE 86 AA; 9433 MW; A9D97A98574E83BE CRC64;

Query Match 11.5%; Score 392; DB 6; Length 86;  
 Best Local Similarity 87.1%; Pred. No. 2e-22;  
 Matches 74; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

107 YICEVEQKEVQLVGLTNSDTHLLOGSLTLTLSPGSSPSVQCRSPRKNIGG 166  
 1 YICEVEDKKEVELVGLTNSDTHLEGSQSLTLTLSPGTPSVKCRSPRKNIGG 60  
 167 KTLVSQLELDGSGTWTCTVLOQNK 191  
 61 KTLVSQLELDGSGTWTCTVLOQNK 85

RESULT 43  
 096GAG  
 ID 096GAG PRELIMINARY; PRT; 614 AA.  
 AC 096GAG;  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=B-cell;  
 RA Strauberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC009851; AAH09851.1; -.  
 DR PIR; S15590; S15590.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR000005; HTHAAC.  
 DR InterPro; IPR007110; IG-like.



```

DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PR00047; IgV_5.
DR SMART; SM00406; IgV_1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
DR PROSITE; PS00835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
DR Hypothetical protein
SQ SEQUENCE 614 AA; 67921 MW; 55EF536E77AA9BBB CRC64;

Query Match
Beet Local Similarity 11.4%; Score 389; DB 4; Length 614;
Matches 150; Conservative 116; Mismatches 272; Indels 144; Gaps 26;

QY 11 LVLQALLPAATQGNKVVLG---KKGTVELTCTAAGSKSI--QFHKNSNQKIL-- 62
DB 7 ILFLVAAATDAVSQMLVQSGAEVKKTGSSVSKASGAYTFTYRYLHWVRAQFQALQEW 66
QY 63 -----GNQGSFLTKGSPSKLNDRAISRSLMDQGNFLITKNIKIEDSDTYICEVEDQ 114
DB 67 MGVITPFGNN-----TNYAQKQODRVTTRDR-SMNTAYMELSLRSEDTANYTC----- 115
QY 115 KEEDLVLEGLTAN-SDTHLQGSLLTLTLESPPGSSPS---VQCRSPRGKNIQSGKTL 169
DB 116 -----ARGYSSMMDAFDIWGQGTMTVSSGSAAPFLPLVSC----- 154
QY 170 SVSOLEQDSGTWCTVVLONQKKVEPKIDIVLAFQ-KASSIVYKKEGQVEFSPPLAT 228
DB 155 -----ENSPSDTSSVAAGCLADFLPDSITFSWKYKNNSSISSTRG-----FPSVLR 201
QY 229 VEKLTGSGELMMQAEASSSSKSMITFDLKNKEVSVKRVYQDDKLQMGKKPLHL--TLRQ 286
DB 202 GGVYAATSOVLLPSKOV-----MGCTDEHVCKVQHPNGNKEKNVPLPVIAELRP 251
QY 287 ALPOYA-----GSGNLTLEAKTKGKHQEVNLVNR-----ATOLQKNTLCEVW 331
DB 252 KVSIVFVRPRDGFPGNPRKSKLTCQATGSPRQIYWSLMEKGQVSGVTTDQVQAEAKS 311
QY 332 GFTSPFLMSTLKENKEAVSKREKPVVYLNPEAGMOCGLSDSGQVLE-SNIKVLPTW 390
DB 312 GFTYKVNSTLTIKESD-----WL---SOSMFTCRVDRHGLTFOQNASMSGVDPQ 358
QY 391 STPVEPKSCDKTHTCPRCAPALLGGPSVFLRPKPKDMLSRPEVNCVVVDVSHEDP 450
DB 359 DTAIR-----VFALRPS-FAISFLTKSKLCLVLTLDLTYVD- 393
QY 451 EVKFMVYVGVENVNAKTKPREEOVNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAP 510
DB 394 SVTISGTRNGEAVKTHNTISSEHPATISAVGEASICEEDMNSGERFTCTYTHDLPSP 453
QY 511 IETKTSKAKQRP-REFQVYTLTPPSRDELTKQVSLTCLVKGPRPSDIANVESNGQP-- 566
DB 454 LKQTSIRPRGVALHRRDVLTLPPARQLMLRESATITCLVTGFSPADVFNQMGQGPUS 513
QY 567 ENNYKTPRVLD--SNGSFELYSKLTVDKSRQGGVFSQVHEHLHHYTHQKSLSP 624
DB 514 PKRYVTSAPMPEPQAGRYFAHSILTVSEEWNTGETYCVVAHEALPRVTERTKDS- 572
QY 625 GLQDETCAEADQDELGLMTT 646
DB 573 ----TEGEVSADBEGFENIMAT 590

RESULT 44
ID 077601 PRELIMINARY; PRT; 86 AA.
AC 077601;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Lophocobus albigena albigena.

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Lophocobus.
OC NCBI_TaxID=75568;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=96320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini)."
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL; AF057391; AAC25135.1; -.
DR HSSP; P01730; 1CDY.
DR GO; GO:0016021; C:Integral to membrane, IEA.
DR InterPro; IPR007110; I9-1-like.
KW Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT NON TER 1 1
FT DOMAIN 1 19 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT DISUPID 49 78 BY SIMILARITY.
FT NON TER 86 86
SQ SEQUENCE 86 AA; 9463 MW; 2BD97A88464FE9AB CRC64;

Query Match
Beet Local Similarity 11.4%; Score 388; DB 6; Length 86;
Matches 74; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICEVEDQKEEYOLVFLGTANSDTHLQGSLLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
DB 1 YICEVEDQKEEYELLVFLGTANSDTHLLEGGSLTLTLESPPGTSVVCGRSPRGKNIQVG 60
QY 167 KTLVSQLELDQSGTWCTVVLONQK 191
DB 61 RTLVSQLELDQSGTWCTVVSQDK 85

RESULT 45
ID 077600 PRELIMINARY; PRT; 86 AA.
AC 077600;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Lophocobus aeternus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Lophocobus.
OC NCBI_TaxID=75566;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=96320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini)."
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL; AF057390; AAC25134.1; -.
DR HSSP; P01730; 1CDY.

```



```

DR GO: GO:0016021; C:Integral to membrane; IEA.
DR InterPro: IPR007110; Ig-like.
KW Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
FT NON_TER 1 1
FT DOMAIN 1 19 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 49 78 BY SIMILARITY.
FT NON_TER 86 86
SQ SEQUENCE 86 AA; 9476 MW; A9D97A88464FE9BE CRC64;

Query Match 11.2%; Score 383; DB 6; Length 86;
Best Local Similarity 85.9%; Pred. No. 9.5e-22;
Matches 73; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 107 YICEVEDKEEVOLVLFGLTANSPTHLQGSLLTLESPPGSSPVQCRSPKGNIOGG 166
DB 1 YICEVEDKEEVOLVLFGLTANSPTHLQGSLLTLESPPGSSPVQCRSPKGNIOVG 60

QY 167 KTLSSVQLELQDSGTWCTVLAQOK 191
DB 61 RTLSSVQLELQDSGTWCTVLAQOK 85

RESULT 46
Q96AA6 PRELIMINARY; PRT; 618 AA.
AC Q96AA6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Lymph;
RA Strauberg R.;
RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC017356; AAH17356.1; -.
DR PIR; S15590; S15590.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00406; IgV_1.
DR PROSITE; PS00835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 618 AA; 67758 MW; 96DBD4C7C69650A6 CRC64;

Query Match 11.2%; Score 383; DB 4; Length 618;
Best Local Similarity 21.8%; Pred. No. 1.6e-20;
Matches 149; Conservative 117; Mismatches 284; Indels 132; Gaps 25;

QY 8 RHLLLVQLALLP-----AATGKRVVLTGKGGDTVELTCTASQKSIQFHWKNSNOIKI 61
DB 2 KHLMFLLVAAPRWVLSQVLOQWAGLLKPSSTLSLTGCVVGGSPGYWMSWIRPPG 61

QY 62 LGNQ--GSFLTKGPKSKNDADRSRL--WDQGNFPLIKNLKIEDSDT--YICEVEDQKE 116
DB 62 KGLEWIGINHSSTNYPNLSKSRVTISVDTSKQSLKSSVNAADTAVVYC----- 114

QY 117 EVQLVFLGSLTANSDTHL---LQGSLLTLESPPGSSPS---VQCRSPKGNIOGGKTL 169
DB 115 --AVITSPASPGTGRYGMVDWGGCTVTVSSGSAAPTLPPLVSC----- 158

QY 170 SVSLELQDSGTWCTVLAQOKKVEFKIDIVLAFO--KASSIVKKEGEQVFSPLAFT 228
DB 159 -----ENSPSDTSSVAVGVCLAQDFLPSDITFSWKYKNSDISSTRG-----FPSVLR 205

```

```

QY 229 VEKLTSGELIMQWAEARSSSSKSWITFDLKNKEVSVKRVTDQPKLQWKKLPLHL--TLPQ 286
DB 206 GGAATATSYQLBSSKDV-----MQGTDHVVCKVHPNGNKKENPLPVIAELP 255

QY 287 ALQVYA-----GSGNLTALAEKTKLHQEVNLVNR-----ATQLQKLTGEVW 331
DB 256 KVSFVPPRDGFGNPNPKSKLICOATGFSPRQLOVSWLRGKOVGSGVTTDQVAEKES 315

QY 332 GPTSPKMLSLKLENKEAKVSKREKPVWVLPNPAQMMQCLLSQGYLLE-SNIKVLPTW 390
DB 316 GPTTYKATSLTILKESD-----KL---SQSMFTCVDRHGLTFQGNASMCVPDQ 362

QY 391 STVEPKSCDKHTCTCPCPAPELLGSPVFLPPPKKDTLMISRTPEVTVVDVSHEDP 450
DB 363 DTAIR-----VPAIPDS--FASIFLTSTKLTCLVLTDLTYD- 397

QY 451 EYKFNWYVDGEVHNNAKTRREQYNSTYVAVSLVTLHODMLNGKRYCKVSKALPAP 510
DB 398 SVTISWTRNGEAVKTHNTNISESHPNATFSAVGBASICEDDMNSGERFTCTVHTDLPSP 457

QY 511 IEKTSKAKGQP--REPQVYTLPPSRDEL-T-KNQVSLTCLVKGFPSPDIAYMESNGQP-- 566
DB 458 LKQTSRPRKVALHRPQVYLLPPAREQLNRESATITCLVTGSPAPVQWQMGQPLS 517

QY 567 ENNYKTPPVLD--SDGSFPLYSKLTVDKSRMOQGNVSCSWHEALHNHYTKSLSP 624
DB 518 PEKVYTSAPWPEQAPGRYPFAHSILTVSEEMWTGETYCVVAHEALPNHVTERTVKS- 576

QY 625 GLQDTERCAEAOGEGLDGLWTT 646
DB 577 -----TSEVSADDEGFENLMAT 594

RESULT 47
Q9BUI0 PRELIMINARY; PRT; 597 AA.
AC Q9BUI0;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Lymph;
RA Strauberg R.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC002963; AAH02963.1; -.
DR HSSP; P01825; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00406; IgV_1.
DR PROSITE; PS00835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 597 AA; 65274 MW; 2DAFA8FB7E055851 CRC64;

Query Match 11.1%; Score 378.5; DB 4; Length 597;
Best Local Similarity 21.9%; Pred. No. 3.3e-20;
Matches 145; Conservative 115; Mismatches 274; Indels 127; Gaps 24;

QY 8 RHLLLVQLALLP-----AATGKRVVLTGKGGDTVELTCTASQKSIQFHWKNSNOIKI 61
DB 2 KHLMFLLVAAPRWVLSQVLOQWAGLLKPSSTLSLTGCVVGGSPGYWMSWIRPPG 61

QY 62 LGNQ--GSFLTKGPKSKNDADRSRL--WDQGNFPLIKNLKIEDSDT--YICEVEDQKE 116
DB 62 KGLEWIGINHSSTNYPNLSKSRVTISVDTSKQSLKSSVNAADTAVVYC----- 114

```



```

Qy 117 EVQLLVFGLTANS DTHL-----LOGQSLTLTLSPSPSSPS-----VQCRSPRGKNIQGGKTL 169
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 115 --ARVITRASPGRDGRGMDVWGQGTITVTVSSGSAAPLFLPVLC----- 158
Qy 170 SVSQELDQSGTWTCTVYLNQKKEKVIDIVLAQ--KASSIVYKKEGQVEFPPLAFT 228
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 159 -----ENSPSDTSSVAVGCLAQDFLPDSITFSWKYKNSDISSTRG-----FPSVLR 205
Qy 229 VEKLTGSGELMMQAEASSSKSWITFDLNKKEVSVKRYTQDPKLGKPLHL--TLPQ 286
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 206 GGTAAATGQVLLPSKV-----MGTDEHVCKVQHPNGKKEKNVPLPVIAELRP 255
Qy 287 ALPOYA-----GSGNLTALAEAKTKLHGEVNLVNR-----ATOLQKNTLCEVW 331
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 256 KVSVPFRPRDGFPGNRKSKLICQATGSPRQIQVSWLRBKGQVSGVTTDQVQAEAKS 315
Qy 332 GPTSPKMLSLKLENKEAKVSKREKPVWVLPNPAAGMOCLSDSGVILE--SNIKVLPW 390
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 316 GPTTYKVTSTLTIKESD-----WL---SQSMFCRVDRHGLTFQONASSMCVPDQ 362
Qy 391 STPVPKSCDKHTKPCPCAPRLGGSPVFLPRPKKOTLMISRTPEVTCVAVDVSHEDP 450
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 363 DTAIR-----VFALPPS--FASIFLTSTKTLCLVTLDTLTYD- 397
Qy 451 EVKFMVYDGVENHNAKTRPREEQNSTYRVVSVLTVLHQMNLNGEKYCKVSNKALPAP 510
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 398 STLTSTWNGEAVKTHNTISESHNPATSAVGEASICEEDMNSGERFTCTVTHDLPSP 457
Qy 511 IEKTSKAKGQF--REPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIWAVESNGQF-- 566
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 458 LKQTSIRPKGVALLHRPDVYLLPRAEQMLRESAITTLCTVGFSPADVIVQMMQGGQPLS 517
Qy 567 ENNYKTPPVLD--SDGSFPLYSKLTVDKSRMOQGNVSCVMHEALNHTYQKSLSLSP 624
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 518 PEKVTSAPEPQAPGRYFAHSILTVEEEMTGETYTCVAHEALPNRVTERVDSKT 577
Qy 625 G 625
      |
Db 578 G 578

```

## RESULT 48

```

Q9BOB8 ID PRELIMINARY: PRT: 597 AA.
AC Q9BOB8:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle, and Lymph.
RA Strausberg R.;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006180; AAH06180.1; -
DR EMBL; BC001872; AAH01872.1; -
DR HSSP; P01825; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00893; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 597 AA; 65300 MW; 2DAFAD50A6375851 CRC64;

```

Query Match

11.0%; Score 376.5; DB 4; Length 597;

Best Local Similarity 21.9%; Pred. No. 4.7e-20;  
Matches 145; Conservative 115; Mismatches 274; Indels 127; Gaps 24;

```

Qy 8 RHLLVLQALLP-----AATQGNKVLGKGGDVELTCTASQKSLQFHKNSNQKI 61
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2 KHLWFFLLVNAFRAVLSQVQLQDAGILRSELTSLTCGVYGGFSGYQWSWIRQPG 61
Qy 62 LGNQ--GSFLTGPKELNDRADRSRL--WDQGNFLIKNLKIEBSDT--YICEVEDQKE 116
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 62 KGLEWIGELINSGITNNYPSLSKRVITISVDTSKQSLSLSSVNAADPAVYIC----- 114
Qy 117 EVQLLVFGLTANS DTHL-----LOGQSLTLTLSPSPSSPS-----VQCRSPRGKNIQGGKTL 169
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 115 --ARVITRASPGRDGRGMDVWGQGTITVTVSSGSAAPLFLPVLC----- 158
Qy 170 SVSQELDQSGTWTCTVYLNQKKEKVIDIVLAQ--KASSIVYKKEGQVEFPPLAFT 228
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 159 -----ENSPSDTSSVAVGCLAQDFLPDSITFSWKYKNSDISSTRG-----FPSVLR 205
Qy 229 VEKLTGSGELMMQAEASSSKSWITFDLNKKEVSVKRYTQDPKLGKPLHL--TLPQ 286
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 206 GGTAAATGQVLLPSKV-----MGTDEHVCKVQHPNGKKEKNVPLPVIAELRP 255
Qy 287 ALPOYA-----GSGNLTALAEAKTKLHGEVNLVNR-----ATOLQKNTLCEVW 331
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 256 KVSVPFRPRDGFPGNRKSKLICQATGSPRQIQVSWLRBKGQVSGVTTDQVQAEAKS 315
Qy 332 GPTSPKMLSLKLENKEAKVSKREKPVWVLPNPAAGMOCLSDSGVILE--SNIKVLPW 390
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 316 GPTTYKVTSTLTIKESD-----WL---SQSMFCRVDRHGLTFQONASSMCVPDQ 362
Qy 391 STPVPKSCDKHTKPCPCAPRLGGSPVFLPRPKKOTLMISRTPEVTCVAVDVSHEDP 450
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 363 DTAIR-----VFALPPS--FASIFLTSTKTLCLVTLDTLTYD- 397
Qy 451 EVKFMVYDGVENHNAKTRPREEQNSTYRVVSVLTVLHQMNLNGEKYCKVSNKALPAP 510
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 398 STLTSTWNGEAVKTHNTISESHNPATSAVGEASICEEDMNSGERFTCTVTHDLPSP 457
Qy 511 IEKTSKAKGQF--REPOVYTLPPSRDELTKNOVSLTCLVKGFPSPDIWAVESNGQF-- 566
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 458 LKQTSIRPKGVALLHRPDVYLLPRAEQMLRESAITTLCTVGFSPADVIVQMMQGGQPLS 517
Qy 567 ENNYKTPPVLD--SDGSFPLYSKLTVDKSRMOQGNVSCVMHEALNHTYQKSLSLSP 624
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 518 PEKVTSAPEPQAPGRYFAHSILTVEEEMTGETYTCVAHEALPNRVTERVDSKT 577
Qy 625 G 625
      |
Db 578 G 578

```

## RESULT 49

```

Q8MUK4 ID PRELIMINARY: PRT: 588 AA.
AC Q8MUK4:
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph.
RA Strausberg R.;
RL Submitted (Dec-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019235; AAH19235.1; -
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.

```



DR Pfam; PF00047; Ig; 5.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG\_LIKE; 5.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 KW Hypothetical protein\_  
 SO SEQUENCE 588 AA; 64438 MW; F6C0D8AD82B39FD7 CRC64;

Query Match 10.9%; Score 373.5; DB 4; Length 588;  
 Best Local Similarity 21.9%; Pred. No. 7.8e-20;  
 Matches 144; Conservative 115; Mismatches 273; Indels 127; Gaps 24;

QY 8 RHLLVLDLALP-----AATGKAVVLGKGGVETLTCTASQKKSQFWMKSNQIKI 61  
 DB 2 KHLWFFLLVAAPRWVLSQVLOQWAGLILPSESLTLCVYGGSFSGYVMSIROPFG 61  
 QY LGNQ--GSFLTKGSPKLNDRADRSRL--WDQGNPLTIKLNKIEDSDT--YICEVEDQKE 116  
 DB 62 KGLEWIGINHGSGNNYPSLKSRTVTSVDSKQSLKLSNVNADTAVYCC----- 114  
 QY 117 EVOLLVFGLTANSPTHL--LQGSLLTLLESPPGSSPS--VOCRSRPGKNIOGKTL 169  
 DB 115 --ARVITRASPOTDGRYGMVWGQGTVTTVSSGSASAVTLPLVSC----- 158  
 QY 170 SVSGLLEDSGTWTCTVLOQKKVEFKIDIVLAFO--KASSIVYKKESEQVEFSPPLAFT 228  
 DB 159 -----ENSPSDSSSVAVGLADFLPDSITFSKYNKNSDISSTRG-----FPSVLR 205  
 QY 229 VEKLTGSGELMWQERRASSKSWITFDLKNKEVSRYKVTODPKLOMGKCLPLHL--TLPO 286  
 DB 206 GGRKAATSQVLLPSKDV-----MGTDEHYVCVQYHPNKGKENVLPYIAELRP 255  
 QY 287 ALPOVA-----GSGNLTALAEAKTGLHOENVLVNR-----ATLOKMLTCEVW 331  
 DB 256 KVSFVFPDRDGFPGFNGPRKSLICQATGFSRQIQVSMREGKQVGSVTTQVOABAKES 315  
 QY 332 GPTSPKMLSLKLEKKEKAKVSKREPVVNLNPEAGMOCCLSDSGVLLF--SNIKVLPW 390  
 DB 316 GPTTYKVTSLTLIKESD-----WL-----SOSMFCRVDHGLTFQONASSMCVPDQ 362  
 QY 391 STPEPKSCDKHTNCPCPAPELLGGPSVFLPPPKKDTLMISRTPEVTCVVNVSHEDP 450  
 DB 363 DTAIR-----VFALPPS--FASIFLTSTKTLCLVTDLTITYD-- 397  
 QY 451 EVKFNWYVDGVEVHNAKTPREBOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPAP 510  
 DB 398 SVTISWTRNGEAVYKHTNISSEHPNATFSAVGEASICEDDWNGSERFTCTVHTDLPSP 457  
 QY 511 IEKITSKAKGQ--REPOVYTTTPSRDELTKNOVSLTCLVKGFPYSDIAYVESNGQ-- 566  
 DB 458 LKQGISRKGVALHHRPDVYLLPRAEQNLNRESATITCLVTGFSPADVFQVMQROGPLS 517  
 QY 567 ENNYKTTPEVLD--SDGSFPLYSKLTVDKSRMOQGNVSCVWMEALNHNHTOKSLSL 623  
 DB 518 PEKVTYSAPEPQAGRYFAHSILTVSEEMNGETTYTCVAHAELPNRTERTVDKS 576

## RESULT 50

086TT1 PRELIMINARY; PRT; 375 AA.  
 AC 086TT1.  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Human full-length cDNA clone CS00D006Y102 of neuroblastoma of Homo sapiens (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Neuroblastoma;  
 RA Genoscope;

RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Neuroblastoma;  
 RA Li W.B., Gruber C., Jessee J., Polayes D.,  
 RT "Full-length cDNA libraries and normalization."  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BX161420; CAD61894.1; -  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003597; IG cl.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR Pfam; PF00047; Ig; 3.  
 DR SMART; SM00407; IGV; 3.  
 DR PROSITE; PS00835; IG\_LIKE; 3.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 KW Plasmid.  
 SO SEQUENCE 375 AA; 41272 MW; 7ACD1AF4399C5E9E CRC64;

Query Match 10.5%; Score 357.5; DB 4; Length 375;  
 Best Local Similarity 27.8%; Pred. No. 6.8e-19;  
 Matches 84; Conservative 61; Mismatches 110; Indels 47; Gaps 11;

QY 332 GPTSPKMLSLKLEKKEKAKVSKREPVVNLNPEAGMOCCLSDSGVLLSNIKVL--PT 389  
 DB 94 GPTTYKVTSLTLIKESD-----WL-----SOSMFCRVDHGLTFQONASSMCVPDQ 139  
 QY 390 WSPFVEKSCDKHTNCPCPAPELLGGPSVFLPPPKKDTLMISRTPEVTCVVNVSHED 449  
 DB 140 QDTAIR-----VFALPPS--FASIFLTSTKTLCLVTDLTITYD 175  
 QY 450 PEVFNWYVDGVEVHNAKTPREBOYNSTYRVSVLTVLHODMLNGEKYCKVSNKALPA 509  
 DB 176 -SVTISWTRNGEAVYKHTNISSEHPNATFSAVGEASICEDDWNGSERFTCTVHTDLPSP 234  
 QY 510 PIEKITSKAKGQ--REPOVYTTTPSRDELTKNOVSLTCLVKGFPYSDIAYVESNGQ-- 566  
 DB 235 PLKQGISRKGVALHHRPDVYLLPRAEQNLNRESATITCLVTGFSPADVFQVMQROGPL 294  
 QY 567 -ENNYKTTPEVLD--SDGSFPLYSKLTVDKSRMOQGNVSCVWMEALNHNHTOKSLSL 623  
 DB 295 SPEKVTYSAPEPQAGRYFAHSILTVSEEMNGETTYTCVAHAELPNRTERTVDKS 354  
 QY 624 PG 625  
 DB 355 TG 356

## RESULT 51

Q13969 PRELIMINARY; PRT; 71 AA.  
 AC 013969;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-JAN-1999 (TREMBlrel. 09, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE CD4 protein (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zverev V.V., Blinov V.M., Nedopasov S.A.;  
 RT "Splice-mediated insertion of three exons of CD4 mRNA."  
 RT CD4 gene: identification of three exons of CD4 mRNA."  
 RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95407135; PubMed=7676667;  
 RA Zverev V.V., Sidorov A.V., Nedopasov S.A., Malishova V.V.,  
 RA Udalova I.A., Andzhaparidze O.G., Blinov V.M.;  
 RT "Nucleotide sequence of two exons of the human T-lymphocyte CD4



```

RT receptor gene).";
RL VOPR_V1RUB01.40:100-102(1995).
DR EMBL, X87579; CAA60883.1; -.
DR EMBL, S79267; AAB35273.1; -.
DR PIR, I60083; I60082.
DR HSPF, P01730; ICDY.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR007110; IG-like.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Receptor.
FT NON_TER 71 71
SQ SEQUENCE 71 AA; 7844 MW; ASC9DB4816135C86 CRC64;

Query Match
Best Local Similarity 10.5%; Score 357; DB 4; Length 71;
Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NMRGVPFRHLVLVQLALLPAATGKNVVLGKKGDVELTCTASQKKSIOFMKXNSNLIK 60
DB 1 NMRGVPFRHLVLVQLALLPAATGKNVVLGKKGDVELTCTASQKKSIOFMKXNSNLIK 60
QY 61 ILGNQGSFLTK 71
DB 61 ILGNQGSFLTK 71

RESULT 52
Q7TOR1 PRELIMINARY; PRT; 587 AA.
AC Q7TOR1;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_Taxid=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=22318257; PubMed=12454917;
RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stalcup M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[3]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Klein S., Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC056078; AAH56078.1; -.
KW Hypothetical protein.
SQ SEQUENCE 587 AA; 64933 MW; 646B5FD0FA414EBD CRC64;

Query Match
Best Local Similarity 9.8%; Score 333.5; DB 13; Length 587;
Matches 98; Conservative 51; Mismatches 118; Indels 65; Gaps 13;

QY 342 LKENKEAKVSKREKV-----WYLNPEAGMMOCL--SDSGOVL 380
DB 280 LKNGQTTEGVRVEPEVDKRGVATSYLITRKEMDLDT--LYSCVVEHASSGS-LQ 335
QY 381 ESNIVLPTMSTPVEPKSCDTHTCPCPAPELLGGSEVFLPPPKQDTLMISRTPEVTC 440
DB 336 EKNMSKSLIMCDTPITPTSI-----QVITIPPS-LESIFEKKSATLTC 376
QY 441 VVDVSHDEPVEKFNWY-VDGVEVNAKTKRREGQYNS--TYRVSVLTVLHODMLNKE 497
DB 377 LVSNMANSDELRSISWFKSGTQETPLKTELGDALYNDNRITYSVKTTTVCADENNDK- 435
QY 498 YCKKVSNNALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDI 556
DB 436 FVCKVEHTELASMKEVELFKKGEVNTSVVFPFPLELSKREKATITCLVKGPSPEI 495
QY 557 AVEW--ESNGQPENNYKTP-----PYLSDGSFPLYSKLTVDSKRWQGNVPSGVNH 608
DB 496 FVKMLHKNEAVPKQYINTSINDELLPKGQSGKFFLYSLHTIDIKWDADSDSFSCVGH 555
QY 609 EALNHHYQKSLSLSPG-----LQDFTC 632
DB 556 ESLPLQLQRSIDKSGSKPTNVNVSLLVSLDTC 587

RESULT 53
Q90544 PRELIMINARY; PRT; 684 AA.
AC Q90544;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Novel antigen receptor precursor.
OS Ginglymostoma citratum (Nurse shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Galeoidea; Orectolobiformes;
OC Ginglymostomatidae; Ginglymostoma.
OX NCBI_Taxid=7801;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=95183140; PubMed=7877689;
RA Plejnik M.P.;
RA Plejnik A.S., Avila D., Hughes M., Hughes A., McKinney E.C.,
RT "A new antigen receptor gene family that undergoes rearrangement and
RT extensive somatic diversification in sharks.";
RL Nature 374:168-173(1995).
DR EMBL, U18701; AAB48195.1; -.
DR PIR, S60266; S60266.
DR HSPF, P01842; ZMCG.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; IG1; 6.
DR SMART; SM00407; IG1; 4.
DR PROSITE; PS50835; IG_LIKE; 6.
DR PROSITE; PS00290; IG_MHC; 3.
KW Receptor; Signal.
FT SIGNAL 1 18
FT CHAIN 19 684
NOVEL ANTIGEN RECEPTOR.

```



SQ SEQUENCE 684 AA; 75224 MW; 2FF9D2071CDA6DFD CRC64;  
 Query Match 9.7%; Score 330.5; DB 13; Length 684;  
 Best Local Similarity 23.5%; Pred. No. 1.8e-16;  
 Matches 158; Conservative 83; Mismatches 237; Indels 195; Gaps 29;  
 QY 16 LALLPAAIQGNKVNVLGKGDVTELTCTAS--QKSIQIOWH-KNSNQIKILNGSGFLTKG 72  
 DB 145 VSLHSAATEEORAA-----NRFOVLCLISGYYPENIAVSMOKNTYTI-----TSGFAATTS 194  
 QY 73 PSKLNDRADSRRL-----WDOGNFPLIKNLKIEDSDTYICEVEDQKEEVLVRLGL 125  
 DB 195 PVKTSNDPSCASLAKVLEQWSRKS-----YISQVY----- 226  
 QY 126 TANSDTLLQGSLLTLESPPGSSPVQCRSPRKNIQGGKTLVSQLLELDQSGTWTCT 185  
 DB 227 -SHSATSSNQKKEINST-----SEIAVLRLRP-----TYEIRIMDKSATLVICE 268  
 QY 186 VLQONKKEFKIDIVVLAFOKASSIVYKKEGEQVE--FSFPLATTEKLTGSGELMQA 242  
 DB 269 VLSTVSAG-----VVVSMVMVNGKV--ENEGVQMEPTKMSGNOYLTLRLTSSVVEW--- 317  
 QY 243 ERASSSKSMITPFDLKNKEVSVKRYTODPKLQMGKKLPLHLTPQALPOYAGSGLTLALE 302  
 DB 318 -----QSGVEYTCSAKQDOSSTPVVAKTKARVEPTKP-HLRLLPSPPEIQSTS----- 366  
 QY 303 AKTGKLHQLVNLVVMRATQLOKNTLCEVWGPTSPKMLSLKLENKAKVSKREKPVVLN 362  
 DB 367 -----SATLTLCLIRGFYPRDKVSVS----- 385  
 QY 363 PEAGMWQCLSDSGVLLLESNIKVLPTM-----STPVEPKSCDKHTHC--- 405  
 DB 386 -----WQ-----KQDVSVSANVTNPTLLEODLPSTSLNLTLAVEMKSGAK-YTCTAS 434  
 QY 406 -PP-----CPABELLGGPSVFLFPPEPKPDLMISTEPVTCVVDVSHED 449  
 DB 435 HPPSGSTYKRYIRNOKVDCRQTDI---SVSLIKK-PREEIWTQGTATITVEIV---YSD 486  
 QY 450 PE-VKFNMYVDGVEVHNAKTKPREQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALP 508  
 DB 487 LENIKVFQVNGVEKKGKGVETQNPMSGSKSTIVSKLKVMASEMDSGTEVEYCLVEDSELP 546  
 QY 509 APIETKISKAK-GQRPQVYTLTPSRBEL-TKQVNSLTCLVKGYPSPDIAVESNQ- 565  
 DB 547 TPVKASIKKAVSQNHPRKYVLLHPSTBEIDENPATIMCATINHPAIIYVGMWANDTL 606  
 QY 566 PENNYKTPPYLDSPFLYSKLVKDSRMQGVFSCVWHEALH-----NHYTQK 618  
 DB 607 LDGGRITQVDSBKSGSSGFVTDRLKLTAAEMNSDTTYSCLVGHPSLNDLIRSTYKSNKG 666  
 QY 619 SLSLSPGLQDDET 631  
 DB 667 PSSVAVSVVLSPT 679  
 RESULT 54  
 Q29027 PRELIMINARY; PRT; 99 AA.  
 AC Q29027;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DB CD4, allele 1 (Fragment).  
 OS Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OK NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN#1183; TISSUE=Blood;  
 RX MEDLINE=93329116; PubMed=8335933;  
 RA Gustafsson K., Germana S., Sundt T.M., Sachs D.H., Leguern C.;  
 RT "Extensive allelic polymorphism in an exposed region of the

RT miniature, . . . . ."  
 RL J. Immunol. 151:1365-1370(1993).  
 DR EMBL; X65629; CAA46583.1; -.  
 DR PIR; I47131; S21461.  
 DR HSP; P01730; ICDY.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IgV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 FT NON TER 1  
 FT NON TER 99  
 SQ SEQUENCE 99 AA; 11170 MW; 40BF080699CF5DOC CRC64;  
 Query Match 8.9%; Score 305.5; DB 6; Length 99;  
 Best Local Similarity 60.2%; Pred. No. 9.3e-16;  
 Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;  
 QY 32 KGGDTVELTCTASQKSIQIOWHKNNSQIKILNGSGFL-TKGPSEKLNDRADSRRLMDQG 90  
 DB 1 KAGDLAELPCHSSQKKNLPPNWKNSNOTKILGGHGSFWHTASVTELTSLDSKKNMWDHG 60  
 QY 91 NFPLIKNLKIEDSDTYICEVEDQKEEVLVRLGLTAN 128  
 DB 61 SFPLIKNLLEVTDGTYICEVEDKRIEVLVRLGLTAS 98  
 RESULT 55  
 Q27379 PRELIMINARY; PRT; 478 AA.  
 AC Q27379;  
 DT 01-OCT-2003 (Tremblrel. 25, Created)  
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Hypothetical protein DKFZp686K04218 (Fragment).  
 GN DKFZp686K04218.  
 OS Homo sapiens (Human).  
 CC Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OK NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Human rectum tumor;  
 RA Bloecher H., Boecher M., Mewes H.W., Weil B., Amid C., Oanger A.,  
 RA Fobo G., Han M., Wiemann S.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BX538066; CAD97996.1; -.  
 KM Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 478 AA; 51620 MW; 4AFCB541F3217CA1 CRC64;  
 Query Match 8.9%; Score 302.5; DB 4; Length 478;  
 Best Local Similarity 24.0%; Pred. No. 1.5e-14;  
 Matches 132; Conservative 66; Mismatches 174; Indels 179; Gaps 27;  
 QY 171 VSOLELDQSG-----TWCTVLOKQKVEFKIDIVVLAFOKASSIVYKKEGEQVE 220  
 DB 17 LSGVQLQESGPRGLVKSQGLSLTCTV-----SGGSI----- 47  
 QY 221 FSPFLATTEKLTGSGELMQAERASSKSMITPFDLKNKEVSVKRYTODPKLQMGKKLPL 280  
 DB 48 -----GSGDYFW-----SWI-----GRGLEW 67  
 QY 281 HLTLPLQALPOYAGS-----GNLTALAEAKGKLHQEVN-----LVVMRATQLOKN 325  
 DB 68 -----MGYIYSGSYTYNPSLESRLSTIDTSKNQSLNSLTADTAVYFCARGVGLG 122  
 QY 326 LTCEVWG-----PTSPKLM-TSLKLENKAKVSKR-----EKPVVVLNPEA 365  
 DB 123 TARDIWQGVVTVSSASPTSPKVPFLSLDSTQDQGVVVAACLVQGFPPQEPPLSVTWS 182  
 QY 366 GM-----WQCLSDSGQVLLLESNIKVLPTWSTPVEPKS--CD-KHTPC-----PCPA 410



```
Db 183 GQNTARNPPPSQDASGDLTYTSSQLTLPATQCP-DGKSVTCHVKHYTNPSQDVTPCPV 241
Qy 411 PELGGSPVFLPPKPF-----KDTLMISTPVTCTVVDVSHEDPEVKNNWV 458
Db 242 P-----PPPCCHPRSLHRPALEDLLGSEANWLTCTLGL-RDASGATFTWTP 289
Qy 459 DGEVHNAKTKPREEOYNSTYRVSVLTVLHODMLNGKEYCKVSNKALPAPIEKTISKA 518
Db 290 SSGK--SAVQGPERRLCCGYSVSVLPGCAQPMNNGEFTCTAAHPELKTPLTANITKS 347
Qy 519 KQGPPEPVYTLPPSDELTKNQ-VSLTCLVKGFPYSDIAVWESNGQ--PENNYKTPP 575
Db 348 -GNTFPEVHLLPPPESELALNELVTLTCLARGFSKDVLRVLQSGELPREKYLTMAS 406
Qy 576 VLD-SGQ--SFLYSLKLTVDKSRWQGNVPSGVMHEALHNHTOKSLSPG----- 625
Db 407 ROEPPSGTTFATVSTILRVAABDMKKGDTFSCVSHALPLAFTOKTIDRLAGKPTHNV 466
Qy 626 ----LQLDCTC 632
Db 467 SVMMAEVDGTC 477
```

## RESULT 56

```
Q96K68 PRELIMINARY; PRT; 494 AA.
```

```
AC 096K68;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ14473.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isegai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
Wagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
Takahashi M., Chida Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
Ninomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027379; BAB55072.1; -.
DR PIR; S21205; S21205.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ
SEQUENCE 494 AA; 53088 MW; 9A1D7AB5AE64C0E CRC64;
```

```
Query Match 8.7%; Score 298.5; DB 4; Length 494;
Best Local Similarity 21.9%; Pred. No. 3.1e-14;
Matches 140; Conservative 85; Mismatches 204; Indels 209; Gaps 28;
```

```
Qy 30 LGKGGDTVELTCTAS--OKKSIOFHW-----KNSNQIKLNGQSF-----TKGPSKL 76
Db 30 LVKPGGSLRLSCAASLFSFSTYAMNVRQAPEGKLEWVSISRSIDYIYRDSVKGRFTI 89
Qy 77 NNRADSRRLMDQGNPLIIKNIKIEDSDTYICEVDQKEVQVLVFGITANSPTHLLOG 136
Db 90 S-RDNKNKSLVLYQW-----SLKVDVTAVYCCARSCNGAI--CYGFSF-----WG 132
Qy 137 QSLTTLSPGSSPSPVQCRPRGKNIQGGKTLVSQLELQDSGTWTCTVLNQKQVBEK 196
```

```
Db 133 QGTLVTVASASPTSD-----KVFLSLCSTQPDG-----NVV 164
Qy 197 IDIVLAF--QKASIVYKKEGEQVFP-SFLPAFVEKLTGSGELMQAERASSSKSMIT 253
Db 165 IACLQGFPPQBPPLSVTSESQGYTARNFP-----PSQDASGDLY-----TSSQ---- 210
Qy 254 FDLKKKEVSVKRVTDPLQKQMGKULPLHLTLFQALPQYAGSGNLTALAEAKTKLHQEVN 313
Db 211 -----LTLF----- 214
Qy 314 LVWKRATQ--LQKVLTCVEWGPSTPKMLSLKLENKAKVSKREKRVWVLNBEAGMOGL 371
Db 215 -----ATQCLAKSVTCHVKHYTNPSQDVTP----- 241
Qy 372 LSDSQVLLSNIKVLPTWSTVEPKSCDKHTTCTPCAPELLGSPVFLPPKXDTLM 431
Db 242 -----CPVSTPTPT-----SPSTPTSPSCG-HPRSLHRLALED-LT 279
Qy 432 ISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVSVLTVLHOD 491
Db 280 LQSEANLTCTLTGL-RDASGVTFTTTPSSGK--SAVQGPERRDLCCGYSVSVLPGCAEP 336
Qy 492 WLNGKEYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ-VSLTCLVKG 550
Db 337 WNHGKTFCTTAAYPESKTPPLTATLSKS-GNTPRPEVHLLPPSEELALNELVTLTCLARG 395
Qy 551 FYPSDIAVWESNGQ--PENNYKTPPVLD-SGQ--SFLYSLKLTVDKSRWQGNVPSGS 605
Db 396 FSPKVLVLRWLQSGSELPREKLTWASROBPQGTTFATVSTILRVAABDMKKGDTFSCM 455
Qy 606 VNHREALHNHYTOKSLSPG-----LQLDCTC 632
Db 456 VGHREALPLAFTOKTIDRLAGKPTHNVSVMAEVDGTC 493
```

## RESULT 57

```
Q96KX8 PRELIMINARY; PRT; 496 AA.
```

```
AC 096KX8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strusberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC016369; AAH16369.1; -.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ
SEQUENCE 496 AA; 53391 MW; D346929849040D69 CRC64;
```

```
Query Match 8.7%; Score 296; DB 4; Length 496;
Best Local Similarity 23.5%; Pred. No. 4.9e-14;
Matches 133; Conservative 72; Mismatches 205; Indels 156; Gaps 22;
```

```
Qy 124 GITANSPTHLLOGSLTTLSPGSSPSVQ---CRSRGKNIQGGKTLVSQ-----E 175
Db 29 GLVKSET-----LSLTCTVSGGSISSSSYWGWIRQPGKLEWIANITYSGITVYNS 83
Qy 176 LQDSGTWCTCTVLQNKQVEFKIDIVLAFQKASIVYKKEGEQVFPFLAFTVEKLTGS 235
```



```

Db      84  LKSRVITISVDTSKNO-----LSLAKRVSTAAIDTAVY-----FCARHCHYSNSGRTGA 129
QY      236  GEILMWQIERRASSSKSNITFDLKNKEVSVKRVYQDEPKLOMKKLPJHLTLFQALFOYAGSG 295
Db      130  IDYMGQGLVTVSSASPT-----SPKV-----FPLSLCSTDP-----DG 163
QY      296  NLTLAL-----BAKTGLKHQEVNLVYMRATO--LQ 323
Db      164  NVVIACLVGCFRFEQDEPLSTWSESGQGVYAKRFPSPQASGDLVYTTSSQLTPATQCLAG 223
QY      324  KNLTCVEWGEFSPKMLSLKENKEAKVSKREKPVVILNPAGMMQCLLSDSGVLLSEN 363
Db      224  KSVTCHVKHTNPSQDVTVP-----243
QY      384  IKVLPTMSPTFVEKSCDKHTHTCPGCAPELLGSPVLFPPKPKDPTLMISTREPTCVVY 443
Db      244  ---CPVPSPTP-----SPSTPPTPSPPCC-HPRLSLHRPLRLE-LILGSEANLTTLT 293
QY      444  DVSHEDPEVKFNRYVDGVEVHNAKTFPRBEQYNSTYRVVSVLYTLVDHOPMLNGSEYKCKVS 503
Db      294  GL-RDASGVTFMTPESSGK--SAVGQPPRPDLCCGYSVSVLPGCAEPMMNKGITFTCTAA 350
QY      504  NKALPAPIETKITSKANGQRPBPQVYTLPPRSBELTKNO-VSLTCLVKGFTYSIDIAVEMES 562
Db      351  YPESKTPLTATYLTLSK-GNTRFREPVLHPPESELANLNLVTLTCLARFSPKDVLYRMLO 409
QY      563  NGQ--PENNNKTPPYLD--SDG--SFFLVSKITVDKSMOOGANVSSCSVMHEALHNHYTQ 617
Db      410  GSGQLPREKELVTASQBPSPSGCTTTTAVYISILRVAAEDMKKGDITFSCVNGHEALPLAFIQ 469
QY      618  KSLSLSFG-----LQIDETC 632
Db      470  KTIIDRLAGKPTNVNVVAEVDGTC 495

```

RESULT	58
Q8NSK4	
ID	Q8NSK4
AC	Q8NSK4
DT	01-OCT-2002 (TREMBlrel. 22, Created)
DT	01-OCT-2002 (TREMBlrel. 22, last sequence update)
DT	01-OCT-2003 (TREMBlrel. 25, last annotation update)
DE	Hypothetical protein.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI	TaxID=9606;

RN SEQUENCE FROM N.A.  
 RP TISSUE=Blood;  
 RC Strausberg R;  
 RA Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 RL EMBL, BC032249; AAH32249.1; -.  
 DR InterPro: IPR003599; IG.  
 DR InterPro: IPR007110; IG\_1like.  
 DR InterPro: IPR003597; IG\_c1.  
 DR InterPro: IPR003066; IG\_MHC.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; IG\_4.  
 DR SMART: SM00409; IG\_4.  
 DR SMART: SM00407; IGc1; 2.  
 DR SMART: SM00406; IGv; 1.  
 DR PROSITE: PSS0835; IG\_Like; 4.  
 DR PROSITE: PSS00230; IG\_MHC; 1.  
 DR Hypothetical protein.  
 KW SEQUENCE 499 AA; 53376 MW; 9345C89582054F32 CRC64;

```
Query Match      8.6%; Score 295; DB 4; Length 499;  
Best Local Similarity 25.8%; Pred. No. 5,9e-14;  
Matches 133; Conservative 61; Mismatches 202; Indels 120; Gaps 25  
  
QY    219 VERSFPLAFVEKLTG-----SGLMWMQERASSSKSWTFPLKNKEV 261  
      :|||                  :|||
```

```

Db      1  MEFGLSWFLVAILKGVCEVQLVSGGVARPGSL-----RLSCATSGFTFD---DS 51
Qy      262 SVKRVYTOPDKLOMGKLP--LHLTLPOALPOYAGS--GNLTALBAKGTGLHOEVLVLM 3117
Db      52  GASWROAP-----GKLEWVSSINNNGSGSTYNADSVKGFSTIRDAKNLSLYOMNSLRV 1070
Qy      318 RATOLO-----KNLTC-----EYWG-----PTSPKLM--LSLKLENKAK 3560
Db      108 EDTALTYCARDPTRYKCSGGSCGLGYMDVWGKTTYVSSASFSFSPKVFPLSLCSTQPDGN 1670
Qy      351 VSKR-----EKPVWVLNPEAGM-----WOCLLSDSGQVLLBSNIKYLP--TWSTPVE 3959
Db      168 VVIACLVGGFFQPEPLSVTWSSGCGVATARNFPSSODASGDLTYTSSQTLPLPTQCLAGK 2270
Qy      356 PKSCD-KHHTCP-----PCPABELLGGPSVFLFPKPK-----DTLMIS 4330
Db      228 SVTCHVKKHTYNSDQVYWPCEVPSTPPTPSPST--PTPSPSCCHPRLSLRHMLBELLG 2060
Qy      434 RTPEVYCVVVDVSHEDPEVKFPMVYDGVVHNAAKTRPREBOYNSTYRVVSVLTVLHQDWL 4930
Db      287 SEANLTCTLTGL--RNASGVTFTWTPSSGK--SAGVGPPRDLCGCVSVSSVLPGCAEPMN 3430
Qy      494 NGKEVKCKKSNKALPAPIEKTISKAKGQRPQVYTLPPSRDELTKNQ--VSLTCLYKGFY 5550
Db      344 HKKFTCTCAAYPSKSTPLVATLSTKS--GNTRFEVHLLPPSEELANLELVTLCLAGFS 4020
Qy      553 PEDIIVEMESNQ--PENNYKTTTPVLD--SDG--SFFLSKLTIVDKSRMOQGVFSCSYM 6070
Db      403 PPDVIVLRHLQSGDELPREKTYLTWASRQESQGTTHFAVVISILRYAEDMKGDTFSLMGV 4620
Qy      608 HEALTHYHTOKSLSPG-----LQIDETC 632
Db      463 HEALPLAFQKTIIDLAKPETHVNVSVVAAEVDGTC 498

```

RESULT	59		
Q8NCL6			
ID	Q8NCL6	PRELIMINARY;	PRT; 493 AA.
AC	Q8NCL6;		
DT	01-OCT-2002 (TREMBLrel. 22, Created)		
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Hypothetical protein FLJ90170.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxId=9606;		

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue-Mammary gland:  
 RA Tsogai T., Ota T., Nishikawa T., Hayashi K., Otsubi T., Sugiyama T.,  
 RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,  
 RA Yamamoto Y., Wakamatsu A., Nakamura Y., Kojima S., Nagahara K.,  
 RA Masuko Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,  
 RA Hattori A., Okumura K., Iwayanagi T., Nimomiya K.,  
 RT "NEO human cDNA sequencing project."  
 RL Submitted (Mar-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AK074651; EMBL1114.1; -.  
 DR InterPro: IPR003599; IG.  
 DR InterPro: IPR007110; IG-1like.  
 DR InterPro: IPR003587; IG\_c1.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003586; IG\_v.  
 DR Pfam: PF00047; IG\_4.  
 DR SMART, SM00409; IG; 4.  
 DR SMART, SM00407; IGc1; 2.

Query Match	8.6%	Score 292;	DB 4;	Length 493;
DR PROSITE; PS50835; IG_LIKE; 4.				
DR PROSITE; PS00290; IG_MHC; 1.				
KW Hypothetical protein.				
SQ SEQUENCE 493 AA; 53224 MW; 12ECD7E094777101 CRC64;				



Best Local Similarity 23.0%; Pred. No. 9.8e-14;  
Matches 148; Conservative 80; Mismatches 203; Indels 212; Gaps 31;

QY 25 GNKVVVGKGGDTVELTCTAS--OKSIOGHMKXNSN-----IKIIGNOG-----SFLTK 71  
DB 27 GGGVVL--PGSIRLSCAASGRFRFYDMHWARQSGELENVALIWTGTITTSYSDSK 84  
QY 72 GSKLNDRAISRSLMDQGNFPLIIKNLIKEDSDTYICEVEDQKEVOLLVEGLTANSDT 131  
DB 85 GLRTVS-RDNYKNITLY-----LEMKSLGAEDTAAYYC-ARDQ---GYAGVGVF---D 128  
QY 132 HLLQGSLLTLIESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDSDGTWCTYLVQNOK 191  
DB 129 HMGQGLVTVSSASP--TSP-----KVPELSLCSSTQPPDG----- 160  
QY 192 KVEPKIDIVLAF--OKASIVYKKEGEQVF--SPFLATVEKLTGSGELMMQAEASSS 248  
DB 161 --NVVIACLVQGFPOEPLSVTWSGQGYTARNFP-----PSQDASGLY---TTS 208  
QY 249 KSMITPDLKNKEVSVKVTQDPKLGKGLPLHLTLPOALPOYAGSGNLTLLAEAKTKL 308  
DB 209 Q-----LTLF----- 213  
QY 309 HOEVNLVVMRATQ--LQKNLTCEVWGPIPSPKLMLSLKENKAKVSKREKPVVNLNPERG 366  
DB 214 -----ATQCLAGKSVTCHVKHYTNPSQDVTVP----- 240  
QY 367 MMQCLSDSGVULNESNIKVLPTWSPVPEPKSCDKHTPCPCAPELLGSPVFLPPKP 426  
DB 241 -----CPVSTPPPP-----SPSTPPSPSCC--HPRLSLHRPL 274  
QY 427 KDTLMISRTPEVTCVVDVSHEDPEVKFMYVGVGVHNAKTPREEQNSTYRVVSVLT 486  
DB 275 ED-LILGSBANLTCTLTGL-RDASGVTFTWTPSSGR--SAVGPRPRDLGCGYSVSVLP 330  
QY 487 VLDHQLNGKEYKCKVSNKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKNQ-VSLT 545  
DB 331 GCAEPNHNKFTCTCTAAYPESKTPLATLTSKS-GNTPREPVHLLPPSEBALNLELVLT 389  
QY 546 CLVGFYPSDIAVWESNQ--PENNYKTPPYLD-SDG--SFPLYSKLTVDKSMQGN 600  
DB 390 CLARGSPDPVLRWLGSGSELPREKYLTMASRQEPQGTTFPAVYSILRVAEDMKGD 449  
QY 601 VFSCSVMEALHNHYTKSLSPG-----LQDETC 632  
DB 450 TTSCHVGHALPLAFQKTTIDRLAGKPTVNVSVMAEVDGTC 492

RESULT 60  
Q9NPP6 PRELIMINARY; PRT; 416 AA.

AC Q9NPP6; 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
OS Immunoglobulin heavy chain variant (Fragment).  
OC Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Playvine R., Estivill X., Becarceller M., Sumoy L.  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Aufreay C., Amosge W., Ballabio A., Estivill X., Gibson K.,  
RA Lehrach H., Poustka A., Lundberg J.,  
RT "The European IMAGE consortium for integrated Molecular analysis of  
human gene transcripts."  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL389978; CAB97534.1; -  
DR HSP; P01789; IMCP.

DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR003596; Ig\_v.  
DR Pfam; PF00047; Ig; 4.  
DR SMART; SM00405; Ig; 1.  
DR PROSITE; PS00835; IG\_LIKE; 3.  
DR PROSITE; PS00290; IG\_MHC; 2.  
FT NON TER 1  
SQ SEQUENCE 416 AA; 44786 MW; 8C41708BB8A84687 CRC64;

Query Match 8.5%; Score 291; DB 4; Length 416;  
Best Local Similarity 28.0%; Pred. No. 9.2e-14;  
Matches 103; Conservative 48; Mismatches 137; Indels 80; Gaps 18;

QY 329 EVMG-----PTSPKLM-LSLKENKAKVSKR-----EKPVWNLNPEAGM- 367  
DB 64 DVMGGTIVTVSSASPTSPKFPPLSLDSTPDGNNVVAQVQGFPOEPLSVTWSGSGN 123  
QY 368 -----WQCLSDSGVULNESNIKVLPTWSTVPEPKS--CD-KTHTCP-----PCPABL 413  
DB 124 VTARNFPSPQASGDLTYTSSQLTLPAQCP--DGKSVTCHVKHYTNPSQDVTVP-- 180  
QY 414 LGSPVFLPPKP-----KDTLMISRTPEVTCVVDVSHEDPEVKFMYVNDGV 461  
DB 181 -----PPPCCHPRLSLHRPALBDLLGSEANLTCTLTGL-RDASGATFTWTPSSG 230  
QY 462 EVMNAKTPREEQNSTYRVVSVLTVLHQMVLNGKEYCKVSNKALPAPIEKTISKAKG 521  
DB 231 K-SAVQGPPEKDLGCGYSVSSVLPGCAQPMNHGTFCTAHPKLTPLTNITKS-GN 287  
QY 522 PREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYPSDIAVWESNQ--PENNYKTPPYLD 578  
DB 288 TPREPVHLLPPSEBALNLELVTLCLARGSPKDVLRWLGSGELPREKLTWASRQE 347  
QY 579 -SDG--SFPLYSKLTVDKSMQGNVFCSVMEALHNHYTKSLSPG----- 625  
DB 348 PSQGTTFPAVYSILRVAEDMKGDPTFCMVGHALPLAFQKTTIDRLAGKPTVNVSV 407  
QY 626 -LQDETC 632  
DB 408 MAEVDGTC 415

RESULT 61  
Q9BRV0 PRELIMINARY; PRT; 500 AA.

AC Q9BRV0; 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
OS Hypothetical protein.  
OC Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Prostate;  
RA Strausberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC005951; AAH05951.1; -  
DR HSP; P01789; IMCP.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR003596; Ig\_v.  
DR Pfam; PF00047; Ig; 4.  
DR SMART; SM00406; Ig; 1.  
DR PROSITE; PS00835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 500 AA; 54154 MW; 0A9BF43F2A3CC6D9 CRC64;

Query Match 8.4%; Score 288; DB 4; Length 500;



Best LocalSimilarity 23.1%; Pred. No. 2e-13;  
 Matches 138; Conservative 68; Mismatches 204; Indels 188; Gaps 27

```

Oy      118 VOLVFGILT-ANSDTHLQGGSLTLTLSPSSSPSVOCSPRGNIOGGKTLVSQLEL 176
Db      7  ILFLVAAATGAGOSQVHLTVOSGALEWMS-----PGASVRASCKT----- 43
Oy      177 QDSGWTCITVLQNGKQKVFKIDIVLAPQKASSIYYKKKEGQEVSRFFLAFTVEKLTSG 236
Db      44 --SG-----YAHHTYSII-----WROAPGCG 63
Oy      237 ELMMQOERASSKSWITPDLNKKEVSVKRVKTODPKLQMGKLLPLHLTLPLQALPYOYAGSN 296
Db      64 LEW-----MGWISSPSDN-----TRFAKRF-----QGR 86
Oy      297 LTLALFATGTLHQV-----NLVWRARATQLOGLNLTC-----EYWG----- 332
Db      87 VTLTDTISTSTVYMWELRSLRSDDTAAVYVCARRYCSYSSQNDYYYYYMDWKGCTTVTVS 146
Oy      333 ---PTSPLKM-LSLKLKNEKAKVSR-----EKPRWVNLPEAGM-----MQCLSD 374
Db      147 SASPTSPKRVFLSLSCSTQPDGNNVVAICLVQGFPEPRLSVTMSBEGQGTARNFPPSDA 206
Oy      375 SGQVLLBSNIKVLP-TWSTPVEPKSCD-KTHTCP-----PCPABELLGGPSVFLFPPK 426
Db      207 SGDLVYTTSSQLTLPLATQCLAGSKVYCHMYHYNTPQSDVTVPCRPVSTPTPTPSPT-PTPT 265
Oy      427 K-----DTLMTSRPEVTCVVVDVHEDPEVKYFNMYVVDGVENHANKTKPR 471
Db      266 SPSCCHPLSLHRRPALIEDLLGSEANLTCYTLTGL-RDASGVYFTFTPSGK--SAVQGR 322
Oy      472 EEQYNSTYRVYSVLTVTHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPPVYLP 531
Db      323 DRDLGCGSVSSVSLGCAEPNHGKFTPTCTAAVPESKTPRLPATLSKS-GNTPRPVHLPL 381
Oy      532 PSRRELTLNQ-VSLTCLVYKGFPSPIAYEWESNGQ--FENNYKTPPLVD-SDG--SFPL 585
Db      382 PPSSELALNELVTLTCLARGSPKPKVLRAWLQGGSELREKLYLTWASROEPPQGTTPPAV 441
Oy      586 YSKLTVDVSRMQCGNVFSGSVWHEALAHNYHTQKSLSPG-----LQLDETC 632
Db      442 TSLRVAALMDKKGDTFSCMGHEALPLAFTQETIDRLAGKPTHTNVSAVAEVDGTC 499

RESULT 62
O96DK0 PRELIMINARY; PRT; 496 AA.
AC O96DK0;
DT 01-DEC-2001 (TREMBLrel). 19, Created)
DT 01-DEC-2001 (TREMBLrel). 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel). 25, Last annotation update)
DE Hypothetical protein FLJ25298.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gastric mucosa;
RA Iehibashi T., Kanehori K., Yosida M., Matanabe S., Ishida S., Ono Y.,
RA Hociuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamaehita H., Chiba Y.,
RA Suzuki Y., Hata R., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,
RA Sugiyama T., Irie R., Otsuga T., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK058027; BAB71633.1;
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGV_1.

```

DR	PROSITE	PS50835; IG_LIKE; 4.
DR	PROSITE	PS00290; IG_MHC; 1.
KM	Hypothetical protein.	
SQL	SEQUENCE	496 AA; 53532 MW; C72EE1E247C86FED CRC64;
Query Match	8.4%; Score 287.5; DB 4;	
Best Local Similarity	23.2%; Pred. No. 2,26-13;	
Matches	136; Conservative 66; Mismatches 193; Indels 191; Gaps 27;	
OY	129 SDTHLLD-GGSLLTTLTSSPPSSPSVOCRSBRKNIQGGKTLVSQLELDGCTWTCTVL	187
DB	19 SQVHLVSGAELKLV-----PSSYKVSVC-----	41
OY	188 QNOKKVEFKDIDVLAFQKASSIYKKEGOVERSPFLAFT-VKLTGSGELMWQABRAS	246
DB	42 -----KASANNFRS-----YATTWKQAGCGGLW-----	66
OY	247 SSKSMITFDLKNKEVSKRYTDPKLMGKKLPLHLTLPOALPOYAGS--GNLTALAEK	304
DB	67 -----MGGIIPNF-----GAPVYQNFOQDRTVISADDS	94
OY	305 TGRKHQEVNLVMPATQLO---KNLT-----CEVWG-----PTSPKLM-L	340
DB	95 TTVVMEIITSLTFEDTAFYYCGRLTYGGSGYYIYQHWGQGLVTIVTSASPTSPKVFPL	154
OY	341 SLKLENEAKSKR-----EKRPVNLNPEAGM-----WQCLSDSGCVLLSNIKV	386
DB	155 SLGCTQPDGMVNLACLVQGFPPQEPPLSVTWSESQGVATANFPSPQASDGLYTTSSQLT	214
OY	387 LP-TWSPVPEPKSD-KTHITCP-----PCPAPFLGSPVFLFPPKPK-----	427
DB	215 LPATQCLAGSKSVTCHVGHYTNPSQDVTVPCEPSPPTPTPSPST-PPTPSPSCCHPRLSLH	273
OY	428 ----DTLMISRTPEVTCVWVDVSHEDPEYFENWVDDEVYHNAKTKREQYNSTYRVS	483
DB	274 RPALEDLLIGSEALUTLTTLGL-RDASGVFTPTWSSGK--SAVQGPDDNDLCGCVSYSS	330
OY	484 VLVTLVHDMVINGEKYKCKVSNKALPADIETIKSKAKQPREPOUYTLPPSRDELTXQ-V	542
DB	331 VLPSCAPPMNHGKFTCTTAAYPESKPTLTATLSKS--GNTPRPPEVHLLPPSEBELALNELV	389
OY	543 SLTGLVGFYPSDIAVWESNGQ--PENNYKTTTPRVLD--SDG--SFPLYSKLTVDKSRMQ	597
DB	390 TLTTLAGFSPKQDVLVWMLQSGQLPPEKYLTMASROEPQGTTFVAVTSLIRVAEDMK	449
OY	598 QGANVSCSVMEALNHNHYTKSLISLSPG-----LQLDETCC 632	
DB	450 KGIPTFSCMVGHGHALPLAFTQKTIDRLAGKPTHVNVSVVMAEVDGTC 495	
RESULT 63		
Q72374	PRELIMINARY; PRT; 492 AA.	
ID	Q72374	
AC	Q72374;	
DT	01-OCT-2003 (Tremblere1.25, Created)	
DT	01-OCT-2003 (Tremblere1.25, Last Sequence Update)	
DT	01-OCT-2003 (Tremblere1.25, Last Annotation Update)	
DE	Hypothetical protein DKFZP686C02218 (Fragment).	
GN	DKFZP686C02218.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Human rectum tumor;	
RA	Bloeker H., Boecher M., Mewes H.W., Weil B., Amid C., Oeanger A.,	
RA	Fobo G., Han M., Wiemann S.,	
RL	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; BX58077; CAD98001.1; -.	
KM	Hypothetical protein.	
SQL	SEQUENCE 492 AA; 53776 MW; 1E7A57560F0CA74B CRC64;	
FT	NON TER 1	



Query Match 8.4%; Score 286; DB 4; Length 492;  
 Best Local Similarity 27.9%; Pred. No. 2.8e-13;  
 Matches 102; Conservative 47; Mismatches 137; Indels 80; Gaps 18;

QY 331 WG-----PTSPKLM-LSLKLENKAVSKR-----EKPVWVLPBAGM--- 367  
 DB 142 WQGGTLVTVSASPTSPKVFPLSLDSTPDQGNVVVACLVGFPFQSPPLSVTWSEGGVNT 201  
 QY 368 ---WQGLSDSGQVLLSNIKVLPTWSTPVEPKS--CD-KHTTCP-----PCPABELLG 415  
 DB 202 AANFPSPDASGDLTYTSSQTLPLPAIQCP-DGKSVTCHVKHYTNPSODVTVCPPV---- 256  
 QY 416 GSVFLFPKP-----KDTLMSRTPEVTVVVDVSHDEPEVKFWMYVDGVEV 463  
 DB 257 -----PPPCCHRLSLHRALEDLLIGSEANLTITGL-RDASGATFTWTSBGK- 307  
 QY 464 HNAKTRPREOYNSTYRVVSVLTVLHQDLNKEVKYCKVSNKALPAPIEKTISKAKGQPR 523  
 DB 308 -SAGQPERDLGGCVSVSVLPGCAQPTHTGETFTCTAHPELKTPLTANITKS-GNTF 365  
 QY 524 EQVYTLPPSRDELTKNQ-VSLTCLVKGYFSPDIAVEMESNQ--PENNYKTTPPVLD-S 579  
 DB 366 REEVHLPPSEELALNELVTLTCLARGPSKPDVLRWLGSGQELPREKYLTMASGROEPS 425  
 QY 580 DG--SFLYSLKLTVDKSRWQGNVFSQVMEALNNHYTKSLSPG-----L 626  
 DB 426 QCTTFATVSLRVAEDMKKGDFTSCWGHLEALPLAFQKTIIDLAKPTHTVNVSVMA 485  
 QY 627 QLDETC 632  
 DB 486 EVDGTC 491

RESULT 64  
 ID Q91WR1 PRELIMINARY; PRT; 488 AA.  
 AC Q91WR1;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN IGH-VJ558 OR A1893585.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA Strubeberg R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC013539; AAI13539.1; -.  
 DR MGI; MGI:96486; Igh-VJ558.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003586; IG\_V.  
 DR Pfam; PF00047; Ig; 4.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KW Hypothetical protein  
 SQ SEQUENCE 488 AA; 52964 MW; F12068460B400B9D CRC64;

Query Match 8.3%; Score 282; DB 11; Length 488;  
 Best Local Similarity 24.5%; Pred. No. 5.6e-13;  
 Matches 126; Conservative 78; Mismatches 198; Indels 112; Gaps 24;

QY 147 PSSSSVQCRSPRGKNI-----QCGKTLG-VSGLDSDSGTWTTVLQNKVVEF 195  
 DB 33 PSASVLSCKA-SGYITIDYVYVWVWQSHKSLWEMIGDINPNVNGT-----SYNCK--F 83  
 QY 196 KDIIVLARGKASIVYKKEGSEVERSPFLATVLEKLTGSGELWQAERASSSKSWITPD 255

DB 84 K-GKATLTVDRSSSIAYWQ-----LNNTLSDSDSAVYCARGPVYVSFSYD 128  
 QY 256 LKN--KEYSVKRYVDPDLQMGKPLH-LTLPPALP-----QYAGSNLTIAL 301  
 DB 129 RGDWVGQGLTVTVAEP---AREPTIPLTFPPALSSDPVIGCLLHDYFPGTMTWTW 184  
 QY 302 EAKTGKHQEVNLVVMRATOLQKULTCFVWGPPTSPKMLSLKLENKAVSKREKPVWVL 361  
 DB 185 -GKSGKDTTYNFPFALASGGRYTMSSQTLTPA-----VE 218  
 QY 362 NPEAGMWQGLSDSGQVLLSNIKVLPTWSTPVEPKSCDKHTGCPCPABELLGPVFL 421  
 DB 219 CPEGSVVCSTVQHDSPVQELVNVCPGICSP-----TTPPSPSCQ---PSLSL 264  
 QY 422 FPPKPKDTLMSRTPEVTVVVDVSHDEP-VKRWYVDGVEVHNAKTRPREOYNST-- 478  
 DB 265 QRPALSD-LLLGSDASITCTINGL--RDPGAVFTW-----EPSTGDAVKKKAVQNSCG 316  
 QY 479 -YRVSVTLTVLHQDLNKEVKYCKVSNKALPAPIEKTISKAKGPREOYTLPPSRDEL 537  
 DB 317 CTSVSSVLPGLAERNNSASFTCTVHPESDT-LTGITAKYTVNTFPQVHLPPSEEL 375  
 QY 538 TKNQ-VSLTCLVKGYFSPDIAVEMESNQ--PENNYKTTPPVLDSDG--SFLYSLKLTV 591  
 DB 376 ALNELVSLTCLVRAFPNKEVLRWLGNEELSPESYLVLEFBLKERGEGATTYLVTSVLRV 435  
 QY 592 DKSRWQGNVFSQVMEALNNHYTKSLSPG 625  
 DB 436 SAEIWKQGDQYSCWGHLEALPMNFQKTIIDLKSG 469

RESULT 65  
 ID Q9UP60 PRELIMINARY; PRT; 384 AA.  
 AC Q9UP60;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE SMC73 protein.  
 GN SMC73.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zheng S., Cao J., Cao W., Cai X., Geng L.;  
 RT "Identification and characterization of SMC73, a gene which is down-regulated in colorectal cancer";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF067420; AAC19365.1; -.  
 DR HSP; P01825; 7FAB.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003597; IG\_C1.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR Pfam; PF00047; Ig; 3.  
 DR SMART; SM00407; IGC1; 2.  
 DR PROSITE; PS50835; IG\_LIKE; 3.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 SQ SEQUENCE 384 AA; 40947 MW; BA7ADC3CA5A9DD48 CRC64;

Query Match 8.2%; Score 281.5; DB 4; Length 384;  
 Best Local Similarity 26.8%; Pred. No. 4.3e-13;  
 Matches 93; Conservative 49; Mismatches 132; Indels 73; Gaps 13;

QY 305 TGKHHQEVNLVVMRATQ--LQKNTLCEVWGPTSPKMLSLKLENKAVSKREKPVWVLN 362  
 DB 91 SGLDLYTSSQTLPLPAIQCAKSVTCHVKHYTNPSODVTVCPPV----- 131  
 QY 363 PEAGMWQGLSDSGQVLLSNIKVLPTWSTPVEPKSCDKHTGCPCPABELLGPVFLF 422  
 DB 132 -----CPVSTPTPTP-----SPSTPTPTSPSC-HPRSLH 161



```
Qy 423 PPKKDTLMIKRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKRREQDYNSTYRVV 482
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 162 RPALED-LLLGSEANLTCTLTGL-RDASGVFTWTPSSGK--SAVQSPERDLGGCVSVS 217
Qy 483 SVLVTVLHODWLNGKRYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ- 541
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 218 SVLPQCAEPWNGKFTCTAAVPSKSTPLTATLSKS-GNTPRPVEVHLPPSEELALNEL 276
Qy 542 VSLTCLVNGFVPSDIAVEMESNGQ--PENNTKTPPVLD-SDG--SFPLYSKLTVDKSRW 596
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 277 VTLTCLARGSFPPKVDLVWMLQSGDELPREKYLTMASRQEPSCGTTTFAVTSILRVAAEDW 336
Qy 597 QOQGNVFCSVNHEALHNHYTOKSLSPG-----LQDDETC 632
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 337 KKGDPTFCMVGHEALPLAFTOKTIDRLAGKPTHNVSVMVAEVDGTC 383

RESULT 66
Q29028
ID Q29028 PRELIMINARY; PRT; 99 AA.
AC Q29028;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
Db CD4, allele 2 (Fragment).
OS Sue scrofa (Fig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=#1183; TISSUE=Blood;
RX MEDLINE=93329116; Pubmed=8335933;
RA "Guatafeson K., Germana S., Sundt T.M., Sachs D.H., Leguern C.;
RT "Extensive allelic polymorphism in an exposed region of the
miniduct. . . .";
RL J. Immunol. 151:1365-1370(1993).
DR EMBL; X65630; CAA46584.1; -.
DR PIR; I47132; S21462.
DR HSSP; P01730; 1CDY.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00406; IGv_1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT SEQUENCE 99 AA; 11390 MW; C22955BEA28318F CRC64;

Query Match 8.2%; Score 280.5; DB 6; Length 99;
Best Local Similarity 56.1%; Pred. No. 7.5e-14;
Matches 55; Conservative 19; Mismatches 23; Indels 1; Gaps 1;

Qy 32 KKGDVETLTCTAGSIOGFHKNKSNQIKILGNGSFLTKGP-SKLANDRADSRSLMDQG 90
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 KAGDLAELPCHSSQKKNLPFSMKNSDQIKILSRHNLHMKASVTELSSLDSKKNMWDHG 60
Qy 91 NFPILIKLKLTEDSPTYICEVEDQKEVQLLVFGITAN 128
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SFPLIKLLEVTDGIYICEVEDKRIEVLVFLRLTAS 98

RESULT 67
Q8WY24
ID Q8WY24 PRELIMINARY; PRT; 497 AA.
AC Q8WY24;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
Db SN66 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
NX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zheng S., Shao X., Cao J., Geng L., Fang Y., Dong Q.;
RT "Identification and characterization of SN66, a Ig-like gene which is
RT down-regulated in colorectal cancer.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF283666; AAL36987.1; -.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003065; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_4.
DR SMART; SM00406; IGv_1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
SO SEQUENCE 497 AA; 53665 MW; F24D08DPA5A6355 CRC64;

Query Match 8.2%; Score 279.5; DB 4; Length 497;
Best Local Similarity 25.2%; Pred. No. 8.9e-13;
Matches 118; Conservative 62; Mismatches 196; Indels 93; Gaps 21;

Qy 229 VEKLTSGGELMQAERASSKSWITFDLKNKEYSVK--RVT--QDPKLMGKPLHLT 283
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 56 VROAPGGGLW-----MGWNNPQYGTNTFAQKFGSLTSRDSINTAYVWVLSLS 106
Qy 284 LPQALPOYAGSGNL-----TLLEAKTGKLGHOENVLVVVRATOLQKN- 325
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 107 TEDSAIFCARNLRCGRGFGYVWFDWGGTLYVTSASAPTPKVPPLSLCSTQPPGNV 166
Qy 326 -LTCEVWG--PTSPKMLSLKENKEKAVSKREKPVVWLNPEAGMOCCLSDSGVLLS 382
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 167 VIACLVGFFPQEP--LSVTWSESGGQVYARNP-----PSQDASGLVYTSSTQLPLA 218
Qy 383 -----NKK-----VLPWTSTPPEPKSCDKHTPCPCPAPELLGGSVF 420
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 219 TQCLAGKSVTCHVGHYTNPSQDVTVPQVPSPTPT-----SBSPPTPSPSC-HPRLS 272
Qy 421 LPPPKDPTLMIKRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKRREQDYNSTYR 480
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 273 LHRPALED-LLLGSEANLTCTLTGL-RDASGVFTWTPSSGK--SAVQSPERDLGGCVSVS 328
Qy 481 SVLVTVLHODWLNGKRYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ 540
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 329 VSSVLPQCAEPWNGKFTCTAAVPSKSTPLTATLPKS-GNTPRPVEVHLPPSEELALNEL 387
Qy 541 Q-VSLTCLVNGFVPSDIAVEMESNGQ--PENNTKTPPVLD-SDG--SFPLYSKLTVDKS 594
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 388 ELVTLTCLARGSFPPKVDLVWMLQSGDELPREKYLTMASRQEPSCGTTTFAVTSILRVAAE 447
Qy 595 RMOQGNVFCSVNHEALHNHYTOKSLSPG-----LQDDETC 632
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 448 DMKKGDPTFCMVGHEALPLAFTOKTIDRLAGKPTHNVSVMVAEVDGTC 496

RESULT 68
Q80Z17
ID Q80Z17 PRELIMINARY; PRT; 487 AA.
AC Q80Z17;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
Db Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strauberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC049143; AAH49143.1; -.
```



DR InterPro; IPR003599; IG\_3.  
DR InterPro; IPR007110; IG\_1like.  
DR InterPro; IPR003597; IG\_c1.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003596; IG\_v.  
DR Pfam; PF00047; IG\_3.  
DR SMART; SM00409; IG\_3.  
DR SMART; SM00407; IGc1; 3.  
DR SMART; SM00406; IGv; 1.  
DR PROSITE; PS50835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 2.  
DR Hypothetical protein.  
QM SEQUENCE 487 AA; 53019 MW; 31E2C893900A4D80 CRC64;

	Query March	8.1%;	Score 276.5;	DB 11;	Length 487;	
	Best Local Similarity	23.5%;	Pred. No. 1.5e-11;			
	Matches 129;	Conservative	80;	Mismatches 216;	Indels 123;	Gaps 24
Oy	121 LVFGLTANSDDTHLLOGOS/LITLLES-----PPGSSPSVQCSPPRGKNIGOGKTSLSVSOL	174				
Dd	1 MYLGINCVFIYVELNAGVOSEVLKEAGGGELVPGSGMKLSCAA-----SGFTF-----	48				
Oy	175 ELQDSGTCTCYLQN-QKKVEKIDIVLAFQASISYYKKGEOVESFFLAIFYEKT	233				
Dd	49 ----SNYMNWVRQSPKGLEWVAIRLSNNVATHYAESVGR-----FTISRDD	95				
Oy	234 GSGELMWO-----AERASSSKSMTTFDLKNKEVSVKRVYTODPKLOMGKLP	279				
Dd	96 SKSSVYLGMNNILRAEDTGIIYCTRGRGYDPMNY-FDVMDAGTIV-TVSEP-----AREPT	149				
Oy	280 LH-LTLPOLP-----OYAGSGNLTALEAKTGKLHQBVLVVMRATOLQXNLT	327				
Dd	150 IYPLTFPOLASDPVIGLICLIHDYFPSPGTMYTM-CGSGKDITTVNFPALASGGRYTS	208				
Oy	328 CEWMGPSTPKMLSLKENKEAKVKREKPVVVLMPEAGMOCLSDSQVLLSNIKYL	387				
Dd	209 SOLTLPA-----VECEGESVSCSVOHDSNPVOELNVNCP	243				
Oy	388 PTWSTPVEPKSCDKHTHCPCPAPELLGGPSVFLPRPKEDTLMTSRTEVTCVVVDVSH	447				
Dd	244 GICSP-----TTTPPPSCQ-----PSLSGLRPALFD-TLGSADSICTINGL--	286				
Oy	448 EDPE-VKRNMYVDGJEVNNAKTKPREBOQNST---YRVSVUTLVHLQDMLNKEYCKYS	503				
Dd	287 RDPEGAIVFTW-----EPSTGXDAVQKKAQONCGCYSVSVYPGCABERNNSASFCTVT	341				
Oy	504 NKALPAPLEKTIISKAKGORBPQVYTLPPSRBELTNQO-VSLTCLYKGFPGPDIAVEMES	562				
Dd	342 HPESDT-LTGITAAKYNTVPFOVHLHPRSSSELALNELVSLTCLVRAFNPEXVLVRWLH	400				
Oy	563 NGQ---PENNYKTPPVLDSDG-SFFLYSKLTVDSRMQGNVFSGSYMBEAHHNHYTQ	617				
Dd	401 GNEBLSPESPYSYLFEELKEPEGEGATTYLVTSLVRAEAIWKQDOQYSCHWGHEALPNNFQ	460				
Oy	618 KSLSLSPG 625					
Dd	461 KTIDRLSG 468					
	RESULT 69					
	099KA4					
ID	099KA4	PRELIMINARY;	PRT:	487 AA.		
AC	099KA4					
DT	01-JUN-2001 (TREMBLrel. 17, Created)					
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)					
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)					
Dd	Hypothetical protein.					
OS	Mus musculus (Mouse).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
NCBI	NCBI_TaxId=10090;					
NP	(1)					
	SEQUENCE FROM N.A.					

RA Strauberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC004786; AA04786.1; -.  
DR HSSP; P01810; 2FEJ.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003596; IG\_v.  
DR Pfam; PF00047; IG; 4.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 2.  
KW Hypothetical protein.  
SQ SEQUENCE 487 AA; 52554 MW; 7DC8E96DB33077B CRC64;

Query Match	8.0%	Score 274.5	DB 11	Length 487
Best Local Similarity	21.8%	Pred. No.2.1e-12		
Matches 142	Conservative 81	Mismatches 220	Indels 207	Gaps 28
QY	1	MNRGVPFPHLLLV---	QALLPATQGNKVLGKGGTVELTCTAS--	OKKSIOFHW-- 53
Db	1	MNFGSLIFLVLVKGVCQCEVQLVESGGGLV---	KPGSLSLSCAASGFTTSSVYMSVWR	57
QY	54	---KNSNQIKILGNQGSFLLTKGSKLNDRADSRRLSDQGNFPLIKLKIEDSDTYIC		109
Db	58	QTPERKLEWATIDSGSY-TYYPDNVKGRFTTISRDN-AKNNLVLMQSHLSSEDTAMYYC		115
QY	110	EVEDQKEVQLVGLTANSPDTHLLOGSLTLTESPPGSSPV-QCSRPGKNIQGGKT		168
Db	116	---ARDMGSPYGGISRFD---YWGQGITTYVSSBARNPFTIPLPLPRA-----		159
QY	169	LSVSLQLSDSGTWTCTVLQNKQKVEFKIDIVLAFOKASSIYVKKGEQVFFSPLAFT		228
Db	160	---LSDDPVI1-----GCLINDYPP-----		176
QY	229	VEKLTGSGELMWAQBRASSSKSMITPDLKXKEVGVKRTQDPKIQMGKCLRHLPLQAL		288
Db	177	---SGTMNVYWGSKGQDIT1-----VNFPPLV		200
QY	289	PQ---YAGSGNLT/-ALEAKTGKLEHENVLVMBRATOLQKNLTGCEWVGPTSPKMLSLKL		344
Db	201	ASGGGYTMSQULTLPAYEC-----		219
QY	345	ENKKAQYKSKREKPVWLVNPEKGMQCLISDSGQVLLSENIKVLPTWSTPVPKSCDKTHT		404
Db	220	---PEEESYKCSYQHDNSNAVQEDIVXC---SGPPPP-----		249
QY	405	CPPCPAPELGSPGVLPKPKQDTLMISRPETVCVVVDVSHEDPEKFMVYGVDEVH		464
Db	250	CPPCP-PSC-HPSLSLORPALPD-LLIGSASLTCTLNGLRNPGAV-FTW-----EPS		299
QY	465	NAKTPREBOYNST---YRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTSKAKQ		521
Db	300	TGXDAVQKAAVQNSCGCYSSVSV.PGCAERWNSGASFCTYTHPESDT-LGTIAKIVN		358
QY	522	PREBOYTLPRSRRELTKNQ-VSLTCLVKGYPEDIANEMWNSNQ---PENNYKTPPV		577
Db	359	TFPPQVHLPRPSSBLALNELVSLTCLVRANPKVLRWRLHGNDELSPESTLVPELKE		418
QY	578	DSDG--SFFLYSKLTVDKSRWQOGNVNCSVMHIALHNHYTOKSLSPG		625
Db	419	PGEGATTVLTVSVLRVSAETWKQGDQYSQVVGHEALPWNFPQKTIIDRLSG		468
RESULT 70				
ID	088650	PRELIMINARY;	PRT;	120 AA.
AC	088650;			
DT	01-NOV-1998 (Tremblrel. 08. Created)			
DT	01-NOV-1998 (Tremblrel. 08. Last sequence update)			
DT	01-OCT-2003 (Tremblrel. 25. Last annotation update)			
DE	T-cell surface glycoprotein CD4 (fragment).			
NN	CD4.			
OS	Marmota monax (Woodchuck).			



OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Scurinae;  
 OC NCBI\_TaxID=9995;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Periphereal blood;  
 RX MEDLINE=20094937; PubMed=10627561;  
 RA Guo J.T., Zhou H., Liu C., Aldrich C., Sapurelli J., Whitaker T.,  
 RA Barasas M.I., Mason W.S., Seeger C.;  
 RT "Apoptosis and regeneration of hepatocytes during recovery from  
 trantient hepatitis virus infections.";  
 RL J. Virol. 74:1495-1505(2000).  
 DR EMBL; AF082497; AAC32621.1; .  
 DR HSSP; P01730; 1WIO.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0006955; P:immune response; IEA.  
 DR InterPro; IPR000973; CD4\_TcAg.  
 DR InterPro; IPR007110; IG-like.  
 DR PRINTS; PR00692; CD4TCANTIGEN.  
 FT NON\_TER 1 1  
 FT TER 120 120  
 SQ SEQUENCE 120 AA; 13623 MW; 44967B512D520195 CRC64;

Query Match 8.0%; Score 273; DB 11; Length 120;  
 Best Local Similarity 64.0%; Pred. No. 3.7e-13;  
 Matches 57; Conservative 8; Mismatches 24; Indels 0; Gaps 0;  
 QY 308 LHQENLVVMAATQOKRLTCEVWGPSPKMLSLKLENKAKVSKRKPPVYNLPEKGM 367  
 DB 1 LHQENLVVMAATQOKRLTCEVWGPSPKMLSLKLENKAKVSKRKPPVYNLPEKGM 60  
 QY 368 WQCLLDSDGQVLLSNIKVLPTWSTPVP 396  
 DB 61 WQCLLDSDGQVLLSNIKVLPTWSTPVP 89

RESULT 71  
 Q8VCX4 PRELIMINARY; PRT; 489 AA.  
 AC Q8VCX4;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypoethetical protein.  
 GN IGH-VJ558 OR A1893585.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon;  
 RA Strauberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; BC018322; AA18322.1; .  
 DR MGD; MGI:96486; Igh-VJ558.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; Ig; 4.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KW Hypoethetical protein.  
 SQ SEQUENCE 489 AA; 53208 MW; CC85B1194DAFEF2C CRC64;

Query Match 8.0%; Score 272.5; DB 11; Length 489;  
 Best Local Similarity 24.5%; Pred. No. 3e-12;  
 Matches 135; Conservative 77; Mismatches 210; Indels 129; Gaps 27;  
 QY 118 VQLVFLGITA--NSDTHLQ--GQSLTTLTLESPGSSPVSQCRSPRGKNIQGGKTLVSQ 174  
 DB 118 VQLVFLGITA--NSDTHLQ--GQSLTTLTLESPGSSPVSQCRSPRGKNIQGGKTLVSQ 174

DB 6 VFLELLSVTAGVHSKVLQDQSGAEV-----KPGASVLSCKA-----SGYTFSDYFI 53  
 QY 175 E--LQDSGT-----W-----TCTVLQOKKVEFKIDIVLAFQKASSIVYKKEGEQVEFS 222  
 DB 54 HMKIKRSGGQLEMTGWNPPSGSISKPKK--FR-DATLTADKSSITTV----- 99  
 QY 223 FPLAFVTEKLTGSGELMWOAER-----ASSKSWITFDLKNKEVSVK-RVTQDPKLOMG 275  
 DB 100 ----MDLSRLTSEDSAVYFCARHEDRGNVDGSLAMFYWGQGLTVTVAEAPRPPT--- 152  
 QY 276 KKLPLHITLQALP-----QYAGSGNLTALAEATGKQHQEVLVMAATQLOK 324  
 DB 153 ----YPLFPQALSSDPVILICLIHDFPSGTMVTV--GASGQDITTVNFPALASGRY 207  
 QY 325 NLTCFVWGPSPKMLSLKLENKAKVSKRKPPVYNLPEKGMQCLLDSDGQVLLSNI 384  
 DB 208 TMSQGLTLP-----VCPGSESVKCGVQDSNVGELNV 242  
 QY 385 KVLPTWSTPVPKSCDKTHTCPPAPABELLGPSEVFLPPPKPDTLMIISRTPEVTCVVD 444  
 DB 243 NCPGICSP-----TTPPPSCQ-----FSLIQRPALPD-LILGSDASITCTLNG 287  
 QY 445 VSHEDPE-VKFNMYVDGVEVHNAKTKPREQVNST---YRVSVTLVTHQDMLNGKRYKC 500  
 DB 288 L--RDPGAVFTW-----EPSTGDAVOKKAVQSCGYSVSVLPCCARWMSGASFCK 340  
 QY 501 KVSNAKLPADIEKTISSAKQOPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPSPDIAYE 559  
 DB 341 TVTHPESDT-LTGITIAVTVYTPFPVYHLLPPSEELALNELVSLTCLVAFAFNKELVLR 399  
 QY 560 WESNGQ--DENNYKTPPVLDSDG--SFLYSKLTVDKSRWQGNVFGSCVWHEALHNH 614  
 DB 400 WLHGNELSPESYLVFEPLEKPGEGATTVLTVSVLRVSAEIMKQGDQYSQVNGHEALPMN 459  
 QY 615 YTKSKLSLSPG 625  
 DB 460 FTQKTDRLSLG 470

RESULT 72  
 Q90WB5 PRELIMINARY; PRT; 482 AA.  
 AC Q90WB5;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE T-cell surface glycoprotein CD4 precursor.  
 GN Ahas placythnucos (Domestic duck).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Anseriformes; Anidae; Anas.  
 OC NCBI\_TaxID=8839;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Spleen;  
 RA Chan S.W.S., Middleton D.L., Lundqvist M., Warr G.W., Higgins D.A.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AF378701; AAK59279.1; .  
 DR InterPro; IPR003599; Ig.  
 DR InterPro; IPR007110; IG-like.  
 DR Pfam; PF00047; Ig; 4.  
 DR SMART; SM00409; IG; 3.  
 DR PROSITE; PS50835; IG\_LIKE; 2.  
 KW Signal.  
 FT SIGNAL 1 29  
 SQ SEQUENCE 482 AA; 54624 MW; BDA0B8E6172AD0B CRC64;

Query Match 8.0%; Score 271.5; DB 13; Length 482;  
 Best Local Similarity 25.0%; Pred. No. 3.5e-12;  
 Matches 111; Conservative 69; Mismatches 169; Indels 95; Gaps 19;  
 QY 8 RHLIVQLALLPAADGNKVVLTGKGDYVELTCTASQKSIQPHMK-----NSNQIK 60  
 DB 12 RAVFVLLQGL--THMAHQOIGVEGKVLNLC---KKGDVTVTWREYDAGSSAIIQ 66



```

Qy      6  ILQNGSSFLTGRSGXKLNDAADSRSLMDQGNFLIKKLIKEDSDPTVICEDQKEVOU  120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      67  ILGK---IFKGAPMSDRSET-----NNSGHKLKVSMLRISDAGTYICEGSDRNSISL  118
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      121  LVEGLTANSDPTTHLLQOOSLTLT-LESPGSSPS-----VQCSRPGKN  162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      119  HAVKLTISNGYFLPEDDELYTMHKSPPSQRFSTILFNSHNSRVLPVUQNETPO---  175
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      163  IQGKTLVSQLELODSGTWTCTVLQONQKVEFKI--DIVLAFQKAS-SIVYKKEGOY  219
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      176  ---KALKVKQLOPTDSGTWICMHSDSPSINENISFVVKVLGFEKTHLBRMYAAVDSTV  232
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      220  EFSEPLAFVYEKLTG-----SGELWQAEARASSKSMITFDLKNKEVSVKRVTDPRLOM  274
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      233  TLEWHLNF---RKIGMKEFFTGQLMWQEGNA-----ITYELDFMNAIDAGELRETK---  280
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      275  GKULPRLHLLPQALP-----QVAGSGLTLALFAKTGKLNQVNLVWVARQ--  321
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      281  -KSGALLREPEMKRSTVEVKIKIKIQLKHSGEYTCOLLVNRRIYQSTELVWQNSANP  339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      322  -----LQKULTCEVAGPPTSPKML-----SLKLENKEAKVSKREKPVVNLPEAGW  368
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      340  PGLPRKGAEWTLTCQVSSPIRNVHLMERVNGTGMQDKSKQOS--ETKVEVKTLAYGMW  397
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      369  QCLLSDSGOVLLESN--IKVLPFW  390
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      398  NCHLMEDNNMKLSLNTYVEAPTW  421
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 73  
Q99LA6  
ID Q99LA6 PRELIMINARY; PRT; 484 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hypochemical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Scuriognathia; Muridae; Murinae; Mus  
OX NCBI\_TaxId=10090;  
RN [1]  
RP  
SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC003495; AA004495.1; -.  
DR PIR: F33932; F33932.  
DR HSP: P01810; 2FBJ.  
DR Interpro: IPR007110; Ig-1like.  
DR Interpro: IPR003006; Ig\_MHC.  
DR Interpro: IPR003596; Ig\_v.  
DR Pfam: PF00047; Ig; 4.  
DR SMART: SM00406; IGV; 1.  
DR PROSITE: PS00845; Ig\_LIKE; 4.  
DR PROSITE: PS00280; Ig\_MHC; 2.  
DR Hypochemical protein.  
QK SSQUNCE 484 AA; 52567 MW; 8EAEAAFP9BCF582FA CRC64;

Query Match 8.0%; Score 271.5; DB 11; Length 484;  
 Best Local Similarity 22.6%; Pred. No. 3.5e-12;  
 Matches 144; Conservative 76; Mismatches 217; Indels 201;

QY 13 VLGLALLPAAVQGNKV-----VLGKQDTVELTCTAQSQKSIQF--HKNSNQIKI 61  
Db 4 VMTLLFLMAAASQSIQAQIQVQSGBELKKPKGETIVKISCSASGYPTTDSMHWKQAPGKG 63  
QY 62 LGNQGSF-LTKGPSKLTNDRADSRSL--WDQGNPPLIIKNKIEDSPYICVEEDQKE 117  
Db 64 LKMGWVNIETGESYVADDFKGRFAFSLETASTHTLQINLKAKEDTATYFCARSDYD 123  
QY 118 VQLVFLGLTANSSTHLQGSQSLTTLTLESPGSSPSVQCRSPKNGIKGGKTLVSQLEIQ 177

```

Db 121 IYAM-----DYWGQSTIVATSSSE-----ARNP-----IYYLPLTE 154
Qy 178 DSGTMTCTYLQNOQKVBEKIDIVLAFQKASSIYKKEGEQVEFPPLAFTVEKLTGSGE 237
Db 155 PA-----LSSDPVII-----GCLIHDPF-----SCTVMN 178
Qy 238 LHMQAERASSSSWMTTFDLKKNKEVSKRVATQDPKLQMGKULPHLTLPOALPOYAGSGNL 297
Db 179 VTM-----GKSGK-----DITVNEPPLASGR-----YTMSSQL 209
Qy 298 TL-ALEAKTGKHOEVNLVMKRAIOLQKHLCEVWGPTSPKMLSLKENKEAKVSKREK 356
Db 210 TLPAYCEPGE-----ESVKCSVGHDSN 231
Qy 357 PIVAVNLPEAGMOCILSDSGVILESNIKVLPTWSTPYERKCDKHTTCCPCCAPBELGG 416
Db 232 PVEELD-----VNSCG-----PTPPPP-----TIPSCQ-----255
Qy 417 PSVFLPEPKPDKDTLMISRPEVTCVVVDVSHEDBEVFNMYVDGVEVHNAKTRPREOYN 476
Db 256 PSLSLGRPALBD-LILGSDASITCTLNGLRNRPBGAV-FTW-----EPSTGDAVQKKAVQ 308
Qy 477 ST---YRVVSVLTVHODWMLNGKEKCKVSKALPAPEKLTISKAKQPREPOVYTLTPS 533
Db 309 NSGCGSVASSVYPCGAERNMNSGASFKCTVTPRE-SGLTLGTIAKYTVNTPPPOVHLLPPP 367
Qy 534 RDELATNQO-VSLTCLVYKGFYPSDIDIVEMESNGO---PENNYKTPRPVLADSDG--SFLYS 587
Db 368 SEELANELLSTLCLVRAFPKCVAYVRLHGNBELSPESYLVFEPLEBEGCATLYLVTS 427
Qy 588 KLTVDKSRMOQGNVFPSCVMHEALHNHYTKQSLSLSPG 625
Db 428 VLEVSALETWKQGDQOYSCWNGEHALPMNFTOKTIDRLSG 465

```

```

QY      588  KLTVTDKSRNQGGANVFSCSVNHEALNNHYTKSLSLSPG 625
         |:::|::|::|::|::|::|::|::|::|::|::|::|
Db      428  VLKRSAEITKQGGDOYSQMVGEALPMNFTOKTIDRLSG 465

```

RESULT	74
Q91207	
ID	Q91207
AC	Q91207;
DT	01-DEC-2001 (TREMBLrel. 19, Created)
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE	Hypochemical protein.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX	NBI TaxId=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Strausberg R.
RL	Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; BC010324; AAH0324.1; -
DR	InterPro; IPR007110; IG_1like.
DR	InterPro; IPR003006; IG_MMC.
DR	InterPro; IPR003596; IG_v.
DR	Pfam; PF00047; Ig; 4.
DR	SMART; SMO0406; IGV; 1.
DR	PROSITE; PSS0835; IG_LIKE; 4.
DR	PROSITE; PSS0290; IG_MMC; 2.
DR	Hypochemical protein.
QO	SEQUENCE 486 AA; 52682 MW; 4FEF835125DA70B CRC64;

Query Match	7.9%;	Score 270.5;	DB 11;	Length 486;
Best Local Similarity	26.9%;	Pred. No. 4.2e-12;		
Matches	98;	Conservative 54;	Mismatches 143;	Indels 69;
				Gaps 15

```
QY 282 LTLPLMP-----QTAGSGLTLALAEATGTLGHEVNLVMMRATOLQALTCVEY 330
Db 153 LTLPLRLSDPVTICLLIHDPFPGSTGNVNW-GSGKDIITTVNPPALSGGGYITMSOL 211
QY 331 WGFSPKMLSLKLENKEAKVSKREKPPVWLNPEAGMOCILSDSGVLTLESNIKVLPTW 390
Db 212 TLPA-----VECPGSGSVKCSYOHDSNAVOELDPVK-----242
```



```

QY 391 STPEPKSCDTHTCPCPAPABELLGSPVFLPPPKDQTLMIISRTPEVTCVVVDVSHEDP 450
D 243 SGPPEP-----CPCEP-PSC--HPSLSQRPALBD-LILGSDASITCLINGLRNPEG 290
QY 451 EVKKNWYVDGVEVNAKTKPREEOYNST--YRVSVLTVLHQMNLNGKEKCKVSKAL 507
D 291 AV-FTW-----EPSTGDAVQKAVQNSCGCYSVSVLPGCAERWNSGASFKCTVTHPES 344
QY 508 PAPIEKTISKAKGQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ- 565
D 345 DT-LTGITAKITVTNTPFPQVHLPPPSBELALNELVSLTCLVRAFPNPKVLYRWLHGNEE 403
QY 566 --PENNYKTTTPPVLDSDG--SFPLYSKLTVDKSRNQGNVSCSMHEALHNHYTQKSLS 621
D 404 LSPESYLVEFPLKEPGEATTVLTYSVLRVGAEFTWKGQDQYSCMVGHEALPMNFTQKTID 463
QY 622 LSPG 625
D 464 RLSG 467

```

## RESULT 75

```

Q91WT1 PRELIMINARY; PRT; 481 AA.
ID Q91WT1
AC 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strauberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013490; AAH13490.1; -.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KM Hypothetical protein.
SQ SEQUENCE 481 AA; 52105 MW; 97DF68D159463F65 CRC64;

```

## Query Match 7.9%; Score 270; DB 11; Length 481;

Best Local Similarity 22.0%; Pred. No. 4,5e-12;  
Matches 139; Conservative 69; Mismatches 213; Indels 212; Gaps 25;

```

QY 15 QALIPAAATGNNKVVLGKGDVETLTAS--QKKSIOFHWNKNQIKILGNQ----- 66
D 20 QVQLQSGPE-----LVKPGASVKISCKASGYTFYSYIHW-----VKQRPQGLVWIGW 69
QY 67 SFLTKGPSKLNDRADSRSLM---DQGNFPLINKLKIEDSDTYICEVEDQKEVQLLVF 123
D 70 IYPGDGNTKYNKEKGTTLTADSSSTAYMFLSLTSEDAVYFC----- 115
QY 124 GLTRANSHTLHQLQSLTLTLSPSSPSVOCSPRGNIGQGTLSVSLQLDQSGTW 183
D 116 -TRGGMAFDYWGQGITLTVSSEPARPT----- 143
QY 184 CTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGEQVEFPLATVEKLTSGELMWQAE 243
D 144 -----IYPLTFPQALSSDPVLIIGLIHDYFP-----SGTMVWTV--- 177
QY 244 RAASSKSWITFDLKNKEVSVRVQDPQLQMGKULPLHLTLPQALPQYAGSGLTTL-ALF 302
D 178 -GKSGK-----DITTVNFPPLASGR-----YTWSSQLTLPAVE 211

```

```

QY 303 AKTGKLGHOEVNLVMBATOLQKNLTCEWGPPTS PKLMLSLKENKAKVSKREKPVWLN 362
D 212 CPBG-----SSVKCSYVHDNSNPVDELN 233
QY 363 PEAGMMQCLSDSGQVLLBSNINVLPTWSTPVPKSCDTHTCPCPAPABELLGSPVFLF 422
D 234 VNC-----FGICSP-TPPPSCQ---PSLSLQ 258
QY 423 PPKDQTLMIISRTPEVTCVVVDVSHEDPE-VKENWYVDGVEVNAKTKPREEOYNST--- 478
D 259 RPALBD-LILGSDASITCLINGL--RDPEGAVFTW-----EPSTGDAVQKAVQNSCGC 310
QY 479 YRVSVLTVLHQMNLNGKEKCKVSKALPAPIEKTISKAKGQPREPOVYTLPPSRDEL 538
D 311 YVSVSVLPGCAERWNSGASFKCTVTHPESDT-LTGITAKITVTNTPFPQVHLPPPSBELA 369
QY 539 KNQ-VSLTCLVKGFYPSDIAVEMESNQ--PENNYKTTTPPVLDSDG--SFPLYSKLTV 592
D 370 LNELVSLTCLVRAFPNPKVLYRWLHGNEELSPESYLVEFPLKEPGEATTVLTYSVLRS 429
QY 593 KSRMOGNVSCSMHEALHNHYTQKSLSLSPG 625
D 430 AEIWKQGDQYSCMVGHEALPMNFTQKTIDRLSG 462

```

## RESULT 76

```

Q8KL172 PRELIMINARY; PRT; 482 AA.
ID Q8KL172
AC Q8KL172;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to expressed sequence A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028249; AAH28249.1; -.
DR PIR; F33932; F33932.
DR PIR; PH1105; PH1105.
DR PIR; PH1108; PH1108.
DR PIR; PH1114; PH1114.
DR PIR; PH1118; PH1118.
DR PIR; PH1119; PH1119.
DR PIR; PH1125; PH1125.
DR PIR; PH1126; PH1126.
DR PIR; PH1128; PH1128.
DR PIR; PH1131; PH1131.
DR PIR; PH1134; PH1134.
DR PIR; PH1139; PH1139.
DR PIR; PH1142; PH1142.
DR PIR; PH1149; PH1149.
DR PIR; PH1150; PH1150.
DR PIR; PH1151; PH1151.
DR PIR; PH1152; PH1152.
DR PIR; PH1153; PH1153.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00407; IGC1; 3.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
SQ SEQUENCE 482 AA; 52121 MW; A06FF083E771D084 CRC64;

```



```

Query Match          7.9%; Score 270; DB 11; Length 482;
Best Local Similarity 22.1%; Pred. No. 4.5e-12;
Matches 136; Conservative 70; Mismatches 209; Indels 200; Gaps 24;

QY 30 LGKKDVELCTAS--QKKSIOFW--KNSNQIKILNQSPFLTKPSKLNDRADSR 84
DB 30 LVKPGASVLTSCASAGYFTSYMMHVVKORPGKLEWIGRIDP--NSGGTKYNEKFKSKA 87
QY 85 SLW---DQGNFLLIKNLKIEDSDTYICEVEDQKEVQLVLGLTANSTHLLQGSQSLT 141
DB 88 TLTVDKPSSTAYWQSLSTSEDSAVYYCTRE-----GDYAMDYWGQGTSTV 133
QY 142 TLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKVFKDIYV 201
DB 134 TVSSSEPARREPT-----IYF 147
QY 202 LAFQKASIVYKKEGQVFPSPPLATVETKLTGSGELMWQARASSKSWITFDLKNKEV 261
DB 148 LTFPQALSSDPVILGCLIHDFP-----SGTMMVTM-----GKSGK----- 183
QY 262 SVKRYTQDPKLMQKGLPLHLPLPOLPOYAGSGNLT--ALEAKTGKHLQEVNLVVMRT 320
DB 184 DITTVNFPPLASGGR-----YTWSQQLTLEAVSCPEG----- 216
QY 321 QLOKNLTCFVWGPTSPKMLSLKENKAIVSKREKPVVNLPEAGMQCLSDSGQVLL 380
DB 217 -----BSVKCSVQHDNSFVQVELVNNC----- 237
QY 381 ESNIKVLPWSTPVEPKSCDHTHTCPCPAPBLGLGSPVFLPPPKDTLMTSRPEVTC 440
DB 238 -----PGICSP--TTPPPSCQ-----PGLSLQRPALPD--LLGSDASITC 276
QY 441 VVVDVSHDEPE-VKFWYVDGVEVHNAKTKPREEQNST--YRVSVTLVHODWLNGK 496
DB 277 TLNGT--RDPEGAVFWM-----EPSTGKAVQKAVQNSCGCVSVLPGCAERKNSGA 329
QY 497 EYKCKVSNALPAPIKTIKSKAGQPREPOVYTLPPSRDELTKNQ--VSLTCLVKGFPYSD 555
DB 330 SKKCTVYTHESDT-LTGTLAKTIVTFFPQVHLPPSEBELALNELVSLTCLVRAFNPKS 388
QY 556 IAVWESNQO---PENNYKTPPVLDSDG--SFFLYSKLTVDKSRWQGNVSCSYMHBA 610
DB 389 VLVRLHGNBELSPESYLVEFPLKEPGBGATTYLVTSVLRVSAELMKQDOYSCWGHBA 448
QY 611 LNNHYTKSLSPG 625
DB 449 LPMNFTOKTIDRLSG 463

RESULT 77
Q9DCD9 PRELIMINARY; PRT; 426 AA.
AC Q9DCD9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Adult male kidney cDNA, RIKEN full-length enriched library,
DE clone:0610041A01, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RC MEDLINE=21085660; Pubmed=1217651;
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Konno S., Yamataka I.,
RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischnann M., Gaasterland T., Giesi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

```

```

RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guelinich S., Hill D., Hofmann W., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Rang B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wymshav-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:686-690(2001).
DR EMBL; AK002875; BAB22422.1; -
DR PIR; F33932; F33932.
DR HSSP; P01810; 2FBU.
DR InterPro; IPR007110; Ig-1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_4.
DR SMART; SM00407; IGC1_2.
DR PROSITE; PS00835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
SQ SEQUENCE 426 AA; 45819 MW; 56E1275BA48FEF81 CRC64;

Query Match          7.9%; Score 269.5; DB 11; Length 426;
Best Local Similarity 26.9%; Pred. No. 4.1e-12;
Matches 98; Conservative 54; Mismatches 143; Indels 69; Gaps 15;

QY 282 LTLPOLP-----QYAGSGNLTLEAKTGKHLQEVNLVVMRATOLQKNLTEV 330
DB 93 LTLPRALSSDPVILGCLIHDFPSTGMVTM--GKSGKDTTVNFPPLASGQGYTMSQL 151
QY 331 WGPSTPKMLSLKENKAIVSKREKPVVNLPEAGMQCLSDSGQVLESNIKVLPTM 390
DB 152 TLPA-----VCCPEGESVYKCVQHDNSAVQELDVXC----- 182
QY 391 STPVEPKSCDHTHTCPCPAPBLGLGSPVFLPPPKDTLMTSRPEVTCVVDVSHDP 450
DB 183 SGPPPP-----CPCCP-PSC--HPSLSLQRPALPD--LLGSDASLTCLNGLRNPEG 230
QY 451 EVKFWYVDGVEVHNAKTKPREEQNST--YRVSVTLVHODWLNGEKYKCKVSNAL 507
DB 231 AV-FTW-----EPSTGKAVQKAVQNSCGCVSVLPGCAERKNSASRKCTVTHES 284
QY 508 PAPIKTIKSKAGQPREPOVYTLPPSRDELTKNQ--VSLTCLVKGFPYSDIAWESNQO- 565
DB 285 DT-LTGTLAKTIVTFFPQVHLPPSEBELALNELVSLTCLVRAFNPKVLRVRLHGNB 343
QY 566 --PENNYKTPPVLDSDG--SFFLYSKLTVDKSRWQGNVSCSYMHBAIINNHYTKSL 621
DB 344 LSPESYLVEFPLKEPGBGATTYLVTSVLRVSAELMKQDOYSCWGHBAI LPMNFTOKTID 403
QY 622 LSPG 625
DB 404 RLSG 407

RESULT 78
Q91WT3 PRELIMINARY; PRT; 481 AA.
AC Q91WT3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN IGH-VJ558 OR A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN
RN SEQUENCE FROM N.A.

```



RC TISSUE=Colon;  
 RA Strausberg R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC013488; AAH13488.1; -  
 DR MGI: 96486; Igh-VJ558.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; Ig; 4.  
 DR SMART: SM00406; IGV; 1.  
 DR PROSITE: PS50835; IG\_LIKE; 4.  
 DR PROSITE: PS00290; IG\_MHC; 2.  
 KM Hypothetical protein.  
 SQ SEQUENCE 481 AA; 52022 MW; 4EEB5C253038B718 CRC64;

Query Match 7.8%; Score 266; DB 11; Length 481;  
 Best Local Similarity 23.2%; Pred. No. 9.1e-12;  
 Matches 121; Conservative 75; Mismatches 187; Indels 138; Gaps 21;

QY 182 WTCTVL-----QNGKVEFKIDIVLAFQKASIVYKKEGQVEFSFPLA---FT- 228  
 DB 3 WSCILFLVATATGVNSQVQLQ-----QPGAEIV--RPGASVXLSCKTSQYTFD 50  
 QY 229 -----VEKLTSGGELMWA-----ERASSKSWITFDLKNKEVS 262  
 DB 51 YMMNWKQRPQGLEWIGAIDPFDSYTSYNOKFKGTTLVPTSSSAYMLSLTSDSA 110  
 QY 263 VKRVTQDPKLQWG-----KKLPLHLTLPOLP-----QYAGS 294  
 DB 111 VYFCARGRDSGGYWGQTTLVSSSEPAEPTIPLFPQALSDPVIIGLIHDYFPS 170  
 QY 295 GNLTLAEAKTGKLEHENVLVMRATQLOKNLTCEVWGPTSPKMLSLKENKAKVSKR 354  
 DB 171 GTMNVTW-GKSGKIDITVNFPPALASGGRYTMSQLTLPA----- 209  
 QY 355 EKPWVVLNPEAGMOCCLSDSGVLLSENIKVLPTWSPVEPKSCDKHTTGPCAPELL 414  
 DB 210 -----VECPGEGSVKCSVQHDSPVQELNVNCPGICSP-----TTPPPSCQ-- 252  
 QY 415 GGPVFLPPPKDITLMSRTPEVTCVVVDVSHEDPE-VKFNMYVDGVEVNAKTPREE 473  
 DB 253 --PSLSIORPALED-LLGSDASITCTLNGL--RDEGAVFTW-----EPSTGDAVQKK 302  
 QY 474 QYNST--YRVVSVLTVLHQMVLNGKEYCKVSNKALPAPIEKTISKAKGPREPQVYTL 530  
 DB 303 AVQNSCGCYSSSVLPGCAERNNSGASFCKVTYHESDT-LTGTLAKTVVNTFPPQVHL 361  
 QY 531 PPSRELTKNQ-VSLTCLVKGFPDIAVEMESNQ---PENNYKTPPVLDSDG--SFF 584  
 DB 362 PPSRELALNELVSLTCLVRAFPKXVLRWLDHGNELSPESYLVFEPLKEPGEATTYL 421  
 QY 585 LYSKLVDSRWQGNVSCSYMEHALNHYTKSLSPG 625  
 DB 422 VTSVLRVSAEITWKQGDQYSQWGHGHALPMNFTQKTIIDRLSG 462

RESULT 79  
 Q91X92 PRELIMINARY; PRT; 482 AA.  
 AC Q91X92;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC01181; AAH1181.1; -  
 DR PIR: F33932; F33932.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_V.  
 DR Pfam: PF00047; Ig; 4.  
 DR SMART: SM00406; IGV; 1.  
 DR PROSITE: PS50835; IG\_LIKE; 4.  
 DR PROSITE: PS00290; IG\_MHC; 2.  
 KM Hypothetical protein.  
 SQ SEQUENCE 482 AA; 51865 MW; 312E01F9C1BC7F3C CRC64;

Query Match 7.8%; Score 266; DB 11; Length 482;  
 Best Local Similarity 22.2%; Pred. No. 9.1e-12;  
 Matches 142; Conservative 71; Mismatches 224; Indels 202; Gaps 26;

QY 10 LLLVLQALLPANTQGNKVLGKK-----DVELTCTAS--QKSIQFMKNSNQIK 60  
 DB 4 LALLCLVTPPSCAL-SQVQLKSGPDLVAPSGSLITCTVSGFALTSVAISWVROPCK 62  
 QY 61 ILNGSGFLTKGSPSKLNDRASSRL-WDQGNFPLIKNLKIEDSDT--YICEVEDQKEE 117  
 DB 63 GLEWLGIVTWGVTNNNSALKSLISKNSKQVFLKNNSLQNDIARYYC-ARDNYYE 121  
 QY 118 VQLVFGELTNSDTHLQGSLLTLESPPGSSPSVQCRSPKNGIQQGKTLVSQLELQ 177  
 DB 122 GAMDYW-----GQGTSTVSSSEPAEPT----- 144  
 QY 178 DSGTWCTVLQNGKVEFKIDIVLAFQKASIVYKKEGQVEFSFPLAFTVEKLTSGSE 237  
 DB 145 -----IYPLTFPQALSDPVIIGLIHDYF-----SGTMN 175  
 QY 238 LMMQAEERASSKSWITFDLKNKEVSVKRVTQDPKLQWKKLPLHLPOLPQYAGSGNL 297  
 DB 176 VTW-----GKSGK-----DITVNFPPALASGGR-----YTMSQL 206  
 QY 298 TL-ALEAKTGKLEHENVLVMRATQLOKNLTCEVWGPTSPKMLSLKENKAKVSKREK 356  
 DB 207 TLPVVECPGEG-----EYKCSVQHDSP 228  
 QY 357 PWWVLNPEAGMOCCLSDSGVLLSENIKVLPTWSPVEPKSCDKHTTGPCAPELLGG 416  
 DB 229 PVQELNVN-----GICSP--TTPPPSCQ---- 253  
 QY 417 PSVFLPPPKDITLMSRTPEVTCVVVDVSHEDPE-VKFNMYVDGVEVNAKTPREEQY 475  
 DB 254 PSLSIORPALED-LLGSDASITCTLNGL--RDEGAVFTW-----EPSTGDAVQKAV 305  
 QY 476 NST--YRVVSVLTVLHQMVLNGKEYCKVSNKALPAPIEKTISKAKGPREPQVYTL 532  
 DB 306 QNSCGCYSSSVLPGCAERNNSGASFCKVTYHESDT-LTGTLAKTVVNTFPPQVHL 364  
 QY 533 SRDELTKNQ-VSLTCLVKGFPDIAVEMESNQ---PENNYKTPPVLDSDG--SFF 586  
 DB 365 PSELALNELVSLTCLVRAFPKXVLRWLDHGNELSPESYLVFEPLKEPGEATTYLV 424  
 QY 587 SKLVDSRWQGNVSCSYMEHALNHYTKSLSPG 625  
 DB 425 SVLRVSAEITWKQGDQYSQWGHGHALPMNFTQKTIIDRLSG 463

RESULT 80  
 Q91WP5 PRELIMINARY; PRT; 479 AA.  
 AC Q91WP5;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;



RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon;  
 RA Strausberg R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC013656; AAH13656.1; -  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; Ig\_4.  
 DR SMART; SM00406; IgV\_1.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS0290; IG\_MHC; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 479 AA; 51603 MW; ECB2D087748564F CRC64;

Query Match 7.8%; Score 265.5; DB 11; Length 479;  
 Best Local Similarity 24.7%; Pred. No. 9.8e-12;  
 Matches 134; Conservative 78; Mismatches 207; Indels 123; Gaps 27;

QY 123 FGLTNSDTHLQGGSLTLT-ES-----PGSSPVSQCRSPRGKNIQGGKTLVSQLEL 176  
 DB 3 FGLTLIFLVLTGKGVQVLESVGGGLVPGSGSLKVSQCA-----SGLTFSNYAM-- 53  
 QY 177 QDSGTWCTVLQN-QKVEFKIDIVLAFQKASSIYKKEGQVEFPFLATVTEKLTGS 235  
 DB 54 ----SW---VRQSPKRLKM-----VAALNSNGNTYSD-----TMKGRFTISRDNK 95  
 QY 236 GELIMQAEPASSSKSNIT-----FDLKNKEYSVKRTQDPKLGKGLPLH-LTL 284  
 DB 96 STLTLQ-----MSSLRSEDITFYCVRGVFDVWAGATAV--TVSESP-----AREPTIYPLTF 147  
 QY 285 PQAALP-----QYAGSNLTALBAKTKLHQEVNLVVMRATOLQKNLTGEVMP 333  
 DB 148 PQAALSDPVIIGCLLHDYFPSSGTWNVTV--GKSGKDIITTVNFPALASGGRYTMSSQLTLP 206  
 QY 334 TSPKMLSLKLENKAKVSKREKPVVNLNPEAGMMQCLSDSGQVLTESNLIKVLPTWSP 393  
 DB 207 A-----VECPGESVYKCSVQHDSNPFQELNVNCPGICSP 241  
 QY 394 VEPKSCDKTHTCPCPAPELLGSPVFLFPKPKDTLMISTRPEVTCVVVDVSHEDP- V 452  
 DB 242 -----TTPPPSCQ-----PSLSLQRPALPD-LILGSDASITCTLNG- -RDPGCA 284  
 QY 453 KFNMYVDGEVHNATKPREEOYNST---YRVVSVLTVLHQDLNMGKEYKCKVSKALPA 509  
 DB 285 VFTW-----EPSTGDAVQKAVQNSCCYSVSLPQCAERWNSGASFKCTVTHESDP 339  
 QY 510 PIEKTIKAKQGPREFQVYTLPPSRDELTKNQ-VSLTCLVKGFPSPDIAMVESNGQ--- 565  
 DB 340 -LTGTLIAKTVNTFPQVHLIPRSEELALNELVSLTCLVRAFNPKGVLYRKLHGNBELS 398  
 QY 566 PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOQGNVSCSMHEALHNHYTOKSLIS 623  
 DB 399 PESYLVEFLKEPREGATTYLVTSVLRSAELWKGDDQYSCMVGEALPMNFTQXTIDRL 458  
 QY 624 PG 625  
 DB 459 SG 460

## RESULT 81

Q8VEAO PRELIMINARY; PRT; 484 AA.  
 AC Q8VEAO;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN IGH-VJ358 OR A1893585.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon;  
 RA Strausberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC019425; AAH19425.1; -  
 DR MGD; MGI:96486; Igh-VJ558.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; Ig\_4.  
 DR SMART; SM00406; IgV\_1.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS0290; IG\_MHC; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 484 AA; 52859 MW; F2522DF5ED6288A6 CRC64;

Query Match 7.7%; Score 264; DB 11; Length 484;  
 Best Local Similarity 26.6%; Pred. No. 1.3e-11;  
 Matches 97; Conservative 53; Mismatches 145; Indels 70; Gaps 14;

QY 282 LTLPAALP-----QYAGSNLTALBAKTKLHQEVNLVVMRATOLQKNLTGEV 330  
 DB 150 LTFPQAALSDPVIIGCLLHDYFPSSGTWNVTV--GKSGKDIITTVNFPALASGGRYTMSSQL 208  
 QY 331 WGPSPKMLSLKLENKAKVSKREKPVVNLNPEAGMMQCLSDSGQVLTESNLIKVLPTW 390  
 DB 209 TLPA-----VECPGESVYKCSVQHDSNPFQELNVNCPGIC 243  
 QY 391 STVPEKSCDKTHTCPCPAPELLGSPVFLFPKPKDTLMISTRPEVTCVVVDVSHEDP 450  
 DB 244 SPP-----TTPPPSCQ-----PSLSLQRPALPD-LILGSDASITCTLNG- -RDP 286  
 QY 451 E-VKFNMYVDGEVHNATKPREEOYNST---YRVVSVLTVLHQDLNMGKEYKCKVSKALPA 506  
 DB 287 EGAVFTW-----EPSTGDAVQKAVQNSCCYSVSLPQCAERWNSGASFKCTVTHPE 341  
 QY 507 LPAPIEKTIKAKQGPREFQVYTLPPSRDELTKNQ-VSLTCLVKGFPSPDIAMVESNGQ 565  
 DB 342 SDF-LTGTLIAKTVNTFPQVHLIPRSEELALNELVSLTCLVRAFNPKGVLYRKLHGN 400  
 QY 566 ---PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOQGNVSCSMHEALHNHYTOKSL 620  
 DB 401 ELSPESYLVEFLKEPREGATTYLVTSVLRSAELWKGDDQYSCMVGEALPMNFTQXTI 460  
 QY 621 SLSPG 625  
 DB 461 DRLSG 465

## RESULT 82

Q8KOF2 PRELIMINARY; PRT; 488 AA.  
 AC Q8KOF2;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Similar to expressed sequence A1893585.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon;  
 RA Strausberg R.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC011703; AAH11703.1; -  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003597; IG\_c1.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; Ig\_4.



DR SMART; SM00407; IG1; 3.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 2.  
SQ SEQUENCE 488 AA; 53127 MW; 0E3B156E15733F0 CRC64;

Query Match 7.7%; Score 264; DB 11; Length 488;  
Best Local Similarity 26.6%; Pred. No. 1.3e-11;  
Matches 97; Conservative 53; Mismatches 145; Indels 70; Gaps 14;

OY 282 LTLQALP-----OYAGSGLTLALAEKTKLHQEVLVWMRAATOLQNLCEV 330  
DB 154 LTFPQALSSDPVLIIGCLIHDFPSGTMNVTW-GKSGDITTVNPPALASGRYTMSSQL 212  
OY 331 WGPSTPKMLSLKLENKAKVSKREKPYVNLNPEAGMOCLLSDSGVLTLESNIKVLPTW 390  
DB 213 TLPA-----VECPGESYKCSVGHDSNPVGLVNNCGIC 247  
OY 391 STVEPKSCDKHTPCPCPAPELLGGPSVFLFPPPKDITLMISRTPEVTCVVDVSHEDP 450  
DB 248 SPP-----TTPPPPSQ---PSLSLQRPALSD-LILGSDASITCTLNGL--RDP 290  
OY 451 E-VKFNWYVDGVNHNATKPREBOYNST---YRVSVLTVLHQDWLNGEKYCKVSNKA 506  
DB 291 EGAVFTW-----EPSTGDAVQKRAVQNSCGCYSVSVLPCCAEKRNMSGASFKCTVTHPE 345  
OY 507 LPAPIEKTISKAKGPPEPOVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ 565  
DB 346 SDF-LTGITIAKVTVNTPFPVHLPPRSEELANELLSTCLVRAPFKVLRMLHNGE 404  
OY 566 ---PENNYKTPPVLDSDG--SFPLYSKLTVDKSRWQGNVFSQVMEALHNHYTQSL 620  
DB 405 ELSPESYLVEFPIKEPGEATTVLTSLVSAEIMKQDQYSCMVGHLEALPMNFTQXTI 464  
OY 621 SLSPG 625  
DB 465 DRLSG 469

## RESULT 83

O91XE1 PRELIMINARY; PRT; 480 AA.  
AC O91XE1;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Colon;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC010738; AAH10798.1; -  
DR InterPro; IPR007110; IG\_Like.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; IG\_4.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 2.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 480 AA; 51936 MW; 20B9234EEF2B41ED CRC64;

Query Match 7.6%; Score 261; DB 11; Length 480;  
Best Local Similarity 26.3%; Pred. No. 2.2e-11;  
Matches 96; Conservative 54; Mismatches 145; Indels 70; Gaps 14;  
OY 282 LTLQALP-----OYAGSGLTLALAEKTKLHQEVLVWMRAATOLQNLCEV 330

DB 146 LTFPQALSSDPVLIIGCLIHDFPSGTMNVTW-GKSGDITTVNPPALASGRYTMSSQL 204  
OY 331 WGPSTPKMLSLKLENKAKVSKREKPYVNLNPEAGMOCLLSDSGVLTLESNIKVLPTW 390  
DB 205 TLPA-----VECPGESYKCSVGHDSNPVGLVNNCGIC 239  
OY 391 STVEPKSCDKHTPCPCPAPELLGGPSVFLFPPPKDITLMISRTPEVTCVVDVSHEDP 450  
DB 240 SPP-----TTPPPPSQ---PSLSLQRPALSD-LILGSDASITCTLNGL--RDP 282  
OY 451 E-VKFNWYVDGVNHNATKPREBOYNST---YRVSVLTVLHQDWLNGEKYCKVSNKA 506  
DB 283 EGAVFTW-----EPSTGDAVQKRAVQNSCGCYSVSVLPCCAEKRNMSGASFKCTVTHPE 337  
OY 507 LPAPIEKTISKAKGPPEPOVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ 565  
DB 338 SDF-LTGITIAKVTVNTPFPVHLPPRSEELANELLSTCLVRAPFKVLRMLHNGE 396  
OY 566 ---PENNYKTPPVLDSDG--SFPLYSKLTVDKSRWQGNVFSQVMEALHNHYTQSL 620  
DB 397 ELSPESYLVEFPIKEPGEATTVLTSLVSAEIMKQDQYSCMVGHLEALPMNFTQXTI 456  
OY 621 SLSPG 625  
DB 457 DRLSG 461

## RESULT 84

O8VCV5 PRELIMINARY; PRT; 481 AA.  
AC O8VCV5;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast tumor;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC018455; AAH18455.1; -  
DR InterPro; IPR007110; IG\_Like.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003596; IG\_V.  
DR Pfam; PF00047; IG\_4.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 2.  
KW Hypothetical protein.  
SQ SEQUENCE 481 AA; 52326 MW; 52B44C5826807143 CRC64;

Query Match 7.5%; Score 256.5; DB 11; Length 481;  
Best Local Similarity 22.9%; Pred. No. 4.8e-11;  
Matches 124; Conservative 82; Mismatches 215; Indels 121; Gaps 25;

OY 120 LLVFGLTANSDTHLQOGSITLTLESP---PSSPSVQCRSPRGKNI-----OGG 166  
DB 6 IFLFLSGTTGVH---SEIQOQSGPELVRPETSVVSCSKAGYSFIDYNIYVWKOSHG 61  
OY 167 KTLG-VSQLELDQSGTMTCTVLDNQKVKVEKIIIVLAFQKASSIYKKEGEQVEFFPL 225  
DB 62 KSLIEWIGYIDPYNGS---SYNQK--FK-GKATLTVDKSSNTAF-----99  
OY 226 AFVYEKLTSGGEIWMQERASSSKSWITFDLKNKEVSK-RVTDPTLQMGKKLPLHLTL 284  
DB 100 -MYLNLTSDSAFYFCAKRWYG-AMPAFWQGTIVTVSAESARNPTI-----YLLTL 150  
OY 285 PQALP-----OYAGSGLTLALAEKTKLHQEVLVWMRAATOLQNLCEVWGP 333



```

Db      151 PALSSDPVILGCLIDHFPFSGTMVWTM-GKSGKIDITTVNFPALASGGRYTMSQLTLP 209
Oy      334 TSPKMLSLKLENKEAKVSKREKPVVYNLPKAGMOCCLSDGQVLLSEINIKVL-PTWST 392
Db      210 A-----VECPGESEVSCSVQHDSPVQELDVNCSGPTPP 244
Oy      393 PVEPKSCDHTHTCPPEAPBELGSPSVFLPPPKKDTLMISRTPEVTCVVNVSHEDPEV 452
Db      245 PL-----TIPSCQ-----PGLSLQRPALPD-LILGSDASTITCLNLRPEGAV 287
Oy      453 KKNWVVDGVNHNAAKTPREEQYNST---YRVSVLTVLHODMLNGKEYKCVSNKALPA 509
Db      288 -FTW-----EPSTGKDAVQKAVQNSCGCVSVSVLPGCAERWNSGASFCTVTHPE-SG 340
Oy      510 PLEKITSKAKGQPREQVYVTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ--- 565
Db      341 TLTGTIAKTVNTFPQVHLPPPEBELNLNELLSTCLVRAFNPKEVLRVWLHGNBELS 400
Oy      566 PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOQGVNFCSCVMHEALNHYTKSLSLG 623
Db      401 PSSYLVFPELKEGEGATTVLTYSVLRSATWKGQDQYSCMVGHEALPMNFTOKTIDRL 460
Oy      624 PG 625
Db      461 SG 462

RESULT 85
O7TMK4 PRELIMINARY; PRT; 479 AA.
AC 07TMK4;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6d; TISSUE=breast tumor;
RX MEDLINE=2388257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Schmen C.M., Schuler G.D.,
RA Altchul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uddin T.B., Toehlyuk S., Carninci P., Franke C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6d; TISSUE=breast tumor;
RA Strauberg R.;
RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055905; AAH55905.1; -
KW Hypothetical protein.
SQ SEQUENCE 479 AA; 52209 MW; 532DED9D46D0AED CRC64;

```

```

Query Match 7.4%; Score 252; DB 11; Length 479;
Best Local Similarity 23.8%; Pred. No. 1e-10;
Matches 110; Conservative 62; Mismatches 167; Indels 124; Gaps 21;

Oy      236 GELW--WQAEARSSKSKITFDLKNKEV-----SYK-RVT---QDPK---LQNGKK 277
Db      49 GDMWNNWVROSPEKLEWVS-QIRNKPNYETETYSDSYKGFITSRDSSKGVYLQNNML 107
Oy      278 LP-----LHLLTPQLP-----QYA 292
Db      108 RPDMDGIYCTVEGMDYGRIGISTVSSSARNPITYLPLPALSSDPVILGCLIDHFP 167
Oy      293 GSGNLTLLAEKTKLHQEVNLVWRATQLOKNLTCEWGCPTSPKMLSLKLENKEAKVS 352
Db      168 PGTMTNVWTM-GKSGKIDITTVNFPALASGGRYTMSQLTLP----- 208
Oy      353 KREKPVVYNLPKAGMOCCLSDGQVLLSEINIKVL-PTWSTPVEPKSCDHTHTCPPEAP 411
Db      209 -----VECPGESEVSCSVQHDSPVQELDVNCSGPTPPPI-----TIPSCQ-- 250
Oy      412 ELIGSPVFLPPPKKDTLMISRTPEVTCVVNVSHEDPEVKFMMVVDGVNHNAAKTKPR 471
Db      251 -----PGLSLQRPALPD-LILGSDASTITCLNLRPEGAV-FTW-----EPSTGKDAVQ 298
Oy      472 EEOYNST---YRVSVLTVLHODMLNGKEYKCVSNKALPAPIEKTISKAKGQPREQVY 528
Db      299 KKAQVNSCGCVSVSVLPGCAERWNSGASFCTVTHPE-SGLTGTIAKTVNTFPQVHL 357
Oy      529 TLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ--PENNYKTPPVLDSDG--S 582
Db      358 LLPPEBELNLNELLSTCLVRAFNPKEVLRVWLHGNBELSPESYLVFPELKEGEGATT 417
Oy      583 FFLYSKLTVDKSRMOQGVNFCSCVMHEALNHYTKSLSLSPG 625
Db      418 YLVTYSVLRSATWKGQDQYSCMVGHEALPMNFTOKTIDRLSG 460

RESULT 86
O8K0Z4 PRELIMINARY; PRT; 480 AA.
ID O8K0Z4
AC O8K0Z4;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to expressed sequence A1893585.
DE Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strauberg R.;
RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC029188; AAH29188.1; -
DR InterPro; IPR003599; IG_1like.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig_4.
DR SMART; SM00409; IG_3.
DR SMART; SM00407; IG_3.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
SQ SEQUENCE 480 AA; 51645 MW; 8690A63C669CDBED CRC64;

```

```

Query Match 7.4%; Score 251.5; DB 11; Length 480;
Best Local Similarity 25.5%; Pred. No. 1.e-10;
Matches 93; Conservative 56; Mismatches 145; Indels 71; Gaps 14;

Oy      282 LTLPPQALP-----QYAGSGNLTLLAEKTKLHQEVNLVWRATQLOKNLTCEV 330

```



```

Db      147 LTLPLALSSDVIITGCLIDHYFPGSGTMVW-GKSGKDIITVNFPPALASGGRYTMSSQL 205
Qy      331 WGPSPKMLSLKLENKAKVSKREKFWVNLNPEAGMMQCLSSGCVLLESNIKVL-PT 389
Db      206 TLPA-----VECPSEESVKCSGVQHDNSNFGQLDVNCSGPT 240
Qy      390 WSTPVEPKSCDKTHTCPCPAPABELLGSPVFLPPKPKDTLMI SRTPEVTCVVVDVSHED 449
Db      241 PPPP-----TIPSCQ-----PSLSLQRALED-LLLGSDASITTLINGLNPE 283
Qy      450 PEVKFNWYVDGEVFNNAKTKREEQYNT---YRVVSVLTVLDHDMLGKEIKCKVSNKA 506
Db      284 GAV-FTW-----EPSTGKDAVQKKAQVNSCCGYSVSVLPGCAERMSGASFCTVTHPE 337
Qy      507 LPAPLEKTIKAKQGPREFQVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ 565
Db      338 -SGTLTGIAKYVNTVPPPOVHLPPSEELALNELSLTCLVRAEPKVELVRAHLHGNE 396
Qy      566 ---PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOCGNVFGSCVMHEALHNHYTQKSL 620
Db      397 ELSPESYLVEPLKPEPGGATTYLVTSVLRVSAETWKGQDQYSCMVGHGHALPMNFTQKTI 456
Qy      621 SLSPG 625
Db      457 DRLSG 461

```

## RESULT 87

```

Q99M22 099M22 PRELIMINARY; PRT; 479 AA.
AC Q99M22;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002091; AA02091.1; -.
DR HSSP; P01810; 2EBJ.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 479 AA; 51992 MW; 768E39A138918892 CRC64;

```

Query Match 7.3%; Score 250.5; DB 11; Length 479;  
 Best Local Similarity 25.5%; Pred. No. 1.4e-10;  
 Matches 93; Conservative 56; Mismatches 145; Indels 71; Gaps 14;

```

Qy      282 LTLPLALP-----QYAGSGNLTLEAKTKLHGVNVLVVMRATQLOKRLTCEV 330
Db      146 LTLPLALSDPVITIGCLIDHYFPGSGTMVW-GKSGKDIITVNFPPALASGGRYTMSSQL 204
Qy      331 WGPSPKMLSLKLENKAKVSKREKFWVNLNPEAGMMQCLSSGCVLLESNIKVL-PT 389
Db      205 TLPA-----VECPSEESVKCSGVQHDNSNFGQLDVNCSGPT 239
Qy      390 WSTPVEPKSCDKTHTCPCPAPABELLGSPVFLPPKPKDTLMI SRTPEVTCVVVDVSHED 449
Db      240 PPPP-----TIPSCQ-----PSLSLQRALED-LLLGSDASITTLINGLNPE 282
Qy      450 PEVKFNWYVDGEVFNNAKTKREEQYNT---YRVVSVLTVLDHDMLGKEIKCKVSNKA 506

```

```

Db      283 GAV-FTW-----EPSTGKDAVQKKAQVNSCCGYSVSVLPGCAERMSGASFCTVTHPE 336
Qy      507 LPAPLEKTIKAKQGPREFQVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ 565
Db      337 -SGTLTGIAKYVNTVPPPOVHLPPSEELALNELSLTCLVRAEPKVELVRAHLHGNE 395
Qy      566 ---PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOCGNVFGSCVMHEALHNHYTQKSL 620
Db      396 ELSPESYLVEPLKPEPGGATTYLVTSVLRVSAETWKGQDQYSCMVGHGHALPMNFTQKTI 455
Qy      621 SLSPG 625
Db      456 DRLSG 460

```

## RESULT 88

```

Q99M67 099M67 PRELIMINARY; PRT; 487 AA.
AC Q99M67;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CD4 protein precursor.
GN CD4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=B2;
RC STRAIN=H.B2; TISSUE=Thymus;
RX MEDLINE=99218434; PubMed=10201936;
RA Koskinen R., Lammi M., Tregaskes C.A., Salomonson J., Young J.R.,
RA Vainio O.;
RT "Cloning and modeling of the first nonmammalian CD4.";
RL J. Immunol. 162:4115-4121(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B2;
RA Koskinen R.H., Salomonson J., Tregaskes C.A., Young J.R.,
RA Goodchild M., Bumstead N., Vainio O.;
RT "The chicken CD4 gene has remained conserved in evolution.";
RL Immunogenetics 0:0-0(2002).
DR EMBL; Y12012; CA072740.1; -.
DR EMBL; AJ401223; CAC82027.1; -.
DR InterPro; IPR007110; IG_1like.
DR Pfam; PF00047; IG_3.
DR PROSITE; PS50835; IG_LIKE; 2.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 487 AA; 54985 MW; B9CBA92CE97F45B CRC64;

```

Query Match 7.0%; Score 240; DB 13; Length 487;  
 Best Local Similarity 25.7%; Pred. No. 8.8e-10;  
 Matches 113; Conservative 61; Mismatches 184; Indels 82; Gaps 20;

```

Qy      10 LLLVQLALLPAATQGNKVVLGKGPVETLCTA-SQKSIQFHMK-----SNQIKILG 63
Db      14 VITVLQGLTPFMAQEQEI-GIAGEKVVLSCKAINNQKGTCTWKYKKEVSTIISFS 72
Qy      64 NQGSFLLKGPSKLNDRADRSRLMDQNPFLIKLKIETSDPYICEVEDQKEVQLLVF 123
Db      73 KAQVFCKKAPM-----TRSELNSNSKLLKYSLSLDAGIYTCCYSPVVISLHVF 125
Qy      124 GLTANSDTHLLQGSQSLTLTL-ESPSSSPS-----VQCRSPGKNIOG 165
Db      126 KLTISNSGHFLTMEDDELTMQNSHSQPHLSIKLFNINNDIVTTEILOEAPQ-KYI-- 182
Qy      166 GKTLVSQLELDQSGTYTCTYVLONQKVEFKI--DIYVLAFOKAS-SIVYKKEEQVQFSS 222
Db      183 ---LKLQKLAIDSGTGMCHVNSPISNINISFDVAVLGEFERLEIIVTVGNTAIISS 239

```



```

Oy 223 FPLATVEVLTLS--SGELWMOAERASSSSWITFDLKNKEVSVKRYTQDPKLOMGKPL 28
Db 240 WRLNFRKIKMGKGFPGKLMMEPOGNATIAELLNFS-----VTHQELHRTYK-SN 288
Oy 281 HL-----TLPPALP-QVAGSGNLTALAEKTKGLHGVNLVYMRATQI----- 322
Db 289 HIMFEISGKTDGTGDVVKIPKQVLNHSGGYKCOLLEINGRTSSVRLVYMQYTAIPAGPL 348
Oy 323 ----QKULTCEWGTSPPLTMLSLEN--KEAKYSKREKVVYVNLNPPA-GMOC-LLS 373
Db 349 SRGKMTLLCOVSGEPLPSNAHLMEVRVNGTQMCKKSKQHEAKVEVNVSAFGLMCHIVE 408
Oy 374 DSG-OVLLESNIKVLPTWST 392
Db 409 DNNKISLNTYVEBAHWNS 428

RESULT 89
P70443
ID P70443 PRELIMINARY; PRT; 90 AA.
AC P70443.
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Mutant T-cell surface glycoprotein CD4 (Fragment).
GN CD4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BR/cdJ;
RA Roewenasser O.A., Fairchild S.P., Tomonari K.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U75219; AAB18133.1; -.
DR HSSP; P05540; ICID.
DR InterPro; IPR007110; Ig-like.
FT NON TER 1
FT 1
SO SEQUENCE 90 AA; 9938 MW; 6528B707928A4665 CRC64;

Query Match 6.9%; Score 237; DB 11; Length 90;
Best Local Similarity 59.2%; Pred. No. 1.3e-10;
Matches 45; Conservative 14; Mismatches 17; Indels 0; Gaps 0

Oy 310 QEVNLVYMRATQIQKULTCEWGTSPKLMLSLKBNKEAKYSKREKPVYVNLNPPAGMIO 369
Db 1 QEVNLVYMRATQIQLNLTITCEWGTSPKRLTLTKOENQEARVSEBQKVQVAPETGLMQ 60
Oy 370 CLSDSGOVLLESNIK 385
Db 61 CLSSEGDVKMDSRIO 76

RESULT 90
O96RW7
ID O96RW7 PRELIMINARY; PRT; 5636 AA.
AC O96RW7.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hemelcincin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Trent J.;
RA "Human hemelcincin gene.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156100; AAK68690.1; -.

```

DR GO; GO:0005727; C:extrachromosomal circular DNA; IEA.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.  
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
DR InterPro; IPR000875; Cecropin.  
DR InterPro; IPR001434; DUF1.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR009017; GFP\_like.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003598; Ig\_c2.  
DR InterPro; IPR000169; SHprot\_acsite.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR02035; VWF\_A.  
DR Pfam; PF00008; EGF; 5.  
DR Pfam; PF00047; Ig; 44.  
DR Pfam; PF00090; tSP; 1; 6.  
DR SMART; SM00179; EGF\_CA; 7.  
DR SMART; SM00408; IGC2; 43.  
DR SMART; SM00209; TSP1; 6.  
DR SMART; SM00327; VMA; 1.  
DR TIGRFAMs; TIGR01451; B\_ant\_repeat; 9.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 5.  
DR PROSITE; PS00268; CECROPIN; 1.  
DR PROSITE; PS01186; EGF\_2; 3.  
DR PROSITE; PS01187; EGF\_CA; 8.  
DR PROSITE; PS050835; IG LIKE; 44.  
DR PROSITE; PSS00639; THIOL\_PROTEASE\_HIS; 1.  
DR PROSITE; PSS0092; TSP1; 6.  
DR EGF-like domain, Immunoglobulin domain.  
SQ SEQUENCE 5636 AA; 613660 MM; F000B319CED7B52C CRR64;

```

Query Match      5.6%; Score 190; DB 4; Length 5636;
Best Local Similarity 19.5%; Pred. No. 0.00019;
Matches 154; Conservative 107; Mismatches 253; Indels 274; Gaps 38;

QY      34 GDVETLICTASQKSKSIQIHFWMKNSNQIILGNQSGFLLTKPSKLANDRADRSRLMDQGNFP 93
DB      807 GSNVTLTCYQGYEPPIITKMRLDNMIFSR--PFSVSSISQLATGA----- 851
QY      94 LIINKLKIEDSDIYICEVEDQ---KEEVOLVYGLTA-----NSDTHLLQGOSLTLTL 143
DB      852 LFIINLWASDQGYITCEANENQFGKIQGETTYTIVGLVAPLIGISPSVAIVIEGQQLTLPC 911
QY      144 ESPPGS-----SPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTC-- 184
DB      912 TLLAGNPIPERRWIKNSAMLLQNDYITVRS-----DGSLLHERVOLQGGGEYTCVAS 963
QY      185 -----TYLQNKVYEFKIDIVYLAFOKAS-----SIYKKEGEVOEF 221
DB      964 NVAGTNNKTTSSVVVHVLPITIOHQOIIISTIEGIPVTLPCKASGNPKPSVIMSKGELIST 1023
QY      222 SFPLAFIVYKTLGS--GELIMQAEBAASSKSMI-----TFDLKXNEY-----SV 263
DB      1024 S-----SAKESAGADGSLYYVSPGESSGEVYCTATNTAGYAKRKVQLTVYVRPRVFGDL 1078
QY      264 KRVYDPEKLQMGKLLPHLTLR---QALP-----QYAGSGNLTIA 300
DB      1079 RGLSQDQKVEISVLAGSEVTLPCGVKSLPRPIITMAETQLISFSPFRHPLTPSGSMKIT 1138
QY      301 LEAKT-----GKLHGEVNLVVMRATOLQK-----NLTCSEWGP 333
DB      1139 -EFTTSDSGMYLCVATNIIAGVNVQAVLVNLVHVPKIQGRPHLKVQVGQRVIDICMNGOT 1197
QY      334 TSPKLTML---SLKLENKEAKVSKREKRPVWV---LNPDAQMOQL---SDSGQVYLE 381
DB      1198 PLPIVITMSKSGSTLVLDGEHHVSNPDGTLISDQATPSDAGIYTCVATNIIAGTDETLIHL 1257
QY      382 ----SNIKVL-PTMSSTVEVERPSCKHTCPPCAP-----ELLG-GPSVFL 421
DB      1258 VQEPPTVEDLEPPNPNTTFOERVANQRIEF-PCPAKGTPEPTIKVLHNGRELVTGEEPIISI 1316

```



[illegible]

Q90524	PRELIMINARY;	PRT;	268 AA.
AC Q90524;			
DT 01-NOV-1996 (TrEMBLrel. 01, Created)			
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE Novel antigen receptor (fragment).			
OS Gangliomastoma cirratum (Nurse shark).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;			
OC Elasmobranchii; Galeomorphi; Galeoidea; Orectolobiformes;			
OC Gangliomastomidae; Gangliomastoma.			
OX NCBI_TaxID=7801;			
RN [1]			
RP SEQUENCE OF 1-240 FROM N.A.			
RC TISSUE=Spleen;			
RX MEDLINE=95183140; PubMed=7877689;			
RA Greenberg A.S., Avila D., Hughes M., Hughes A., McKinney E.C.,			
RA Plajunt M.F.,			
RT "A new antigen receptor gene family that undergoes rearrangement and			
RT extensive somatic diversification in sharks.";			
RL Nature 374:168-173 (1995).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC TISSUE=Spleen;			
RA Greenberg A.S.;			
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.			
DR EMBL; U18680; AAB48352.1; -.			
DR HSSP; P01842; 7FAB.			
DR GO; GO:0004872; F:receptor activity; IEA.			
DR InterPro; IPR007110; Ig_11k.			
DR InterPro; IPR003597; Ig_c1.			
DR InterPro; IPR003006; Ig_MHC.			
DR Pfam; PF00047; Ig_2.			
DR SMART; SM00407; IgC1.1.			
DR PROSITE; PS50835; IG_LIKE.2.			
DR PROSITE; PS00290; IG_MHC.1.			
DR Receptor.			
KW NON TER	268	268	
SQ SEQUENCE	268 AA;	29579 MW;	OCDE524DDE9FD4 CRC64;
Query Match	5.4%;	Score 186;	DB 13; Length 268;
Best Local Similarity	24.3%;	Pred. No. 4,9e-06;	
Matches	65;	Conservative 39;	Mismatches 99; Indels 66; Gaps 10
QY	417 PSVFLEPPPKPKDTLMSRPF-----VTCVVVDVSHEDPEVKFNWYDG-----460		
Db	14 PYVF-----TARVDQTPQEIETKEGSGSLINCVLSDNCALPSTYMRKKSGETNEE65		
QY	461 -----VEVNAKTKPREEGYNSTYRVSVLTITLQDMLNGEKYCKV--SNKALPAP510		
Db	66 TISKGRVYETVNSGKSFSLIND-----LTV-----EDSGTYACKYRRKNWAVDGC113		

```

QY 511 IEXT-----ISKXGGRBQYVTL--PBRDELTKNOVSLTCLVKGFPSPDA 557
D6 114 LEBLMDIYYVGGGTGVTVNPGIPLSPPLVSLSHSATEEQRANGFVOLVCLISGYPENIA
QY 558 VERESNGOP-ENNVKTPPVLBDSDFSFLYSKLTJYDSCRMOQGVFSGSVMEALHNHYT 616
D6 174 VSNQKTKTKITTGFPATTSFVTKTSSNDPFCASLLKVPLOEMSRGVSVCQVSHSATSSNOR 233
QY 617 OKSLSLSPGILDETCAQDGEIDGLW 644
D6 234 KEIRSTS-----ETAVLLRDPVYEIN 255

```

	RESULT	92	
	OBWU38		
ID	OBWU38	PRELIMINARY;	PRT; 573 AA.
AC	OBWU38:		
DT	01-MAR-2002 (TREMBLrel. 20,	Created)	
DT	01-MAR-2002 (TREMBLrel. 20,	Last sequence update)	
DT	01-OCT-2003 (TREMBLrel. 25,	Last annotation update)	
DE	Hypothetical protein.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Tonsil;		
RA	Straubeberg R.;		
RL	Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL; BC021276; AAAH21276.1; -		
DR	PIR; S21205; S21205.		
DR	InterPro; IPR007110; Ig-like.		
DR	InterPro; IPR003006; Ig MHC.		
DR	InterPro; IPR003596; Ig_v.		
DR	Pfam; PF00047; Ig; 4.		
DR	SMART; SMO0406; IGV; 1.		
DR	PROSITE; PS00835; IG_LIKE; 4.		
DR	PROSITE; PS00290; IG_MHC; 2.		
KW	Hypothetical protein.		
SQ	SEQUENCE 573 AA; 62967 MW; FD072344033AC530 CRC64;		
Query March	5.4k;	Score 184.5;	DB 4; Length 573;
Best local similarity	21.0k;	Pred. No. 1.9e-05;	
Matches 151;	Conservative 77;	Mismatches 248;	Indels 243; Gaps 34.
OY	1	MNRGVPFRHLILV---	QLALLPATQGNKVLAKSGDPTVELTGTAS--QKKSIOFHNKN 55
Dd	1	MELGLSMIFLLAILIKGVQEYCEVQLVESGGLV---	QPGRSLRLSCAASFETDDVAMHWVR 57
OY	56	SNQIK-----ILGNQGSF---LTGPSKLNDRADSRRLMLQNGNPFLIIKNIKIDS 104	
Dd	58	QAPKGLEWWSGISMNSIGIVADSVXGRFTIS-RDNAKSLYLQMN-----SLRADDT 110	
OY	105	DTYICEVDQKEEYQVLVFGLTANSPTHLOGQSILTLESPPSSPEV-----QCRRSPR 159	
Dd	111	ALYYICAHGSSSYI-----GYIYGMD--VMGGSTTYTVSSAPFKARPVPFFIISGRHPK 162	
OY	160	GKN-----IQGGKTLVSQGLELDGSGTWTCITVLQNKVKVEFKIDIIVLAFAKASSIVY 212	
Dd	163	DNSPVLVALCLITGYHPTSVTV-----TW-----Y 186	
OY	213	KKEBEQYEFSFP-----LAFYEVKLTGSGSELMWOAE-----RASSEKSWITPLDKNYE 260	
Dd	187	MGTSGSQPFQRTPEFIQRDSYYMTSSQLSTPLQWRQGEYKCVCVHTASKS-----KKE 239	
OY	261	VSVKRVTPDPFLQMGKKLPLHLTLPOALPYAGS---GNLTALAEAKTGKLGHOENVLYVM 317	
Dd	240	IF--RWESPFAQAS-----SVTPAQPOALEGSLAAKTATPAATTTRNTGRGEGE----- 284	
OY	318	RAIQDLQNLTCIEWWGPTSPKLMLSLKLENKEAKYSKRREKPVWVINPEAGMNOCLISDSGO 377	



```

Db 285 -----KKKEKEKEEERE----- 298
Qy 378 VLESNIKVLPTWSTFVEPKSCDKHTTCCPCAPPELLGSPVFLPPPKOTLMISRTPE 437
Db 299 -----TKTPECSHTQP--LG---VYLLTPAVQD-LMLRDKAT 330
Qy 438 VTCVVVDVSHEDPEVKFNMYVDG-VEVHNAKTPEEQNSTYRVSVLTVJHQMPLNGK 496
Db 331 FTCFVVGSLDKDAHL--TWEVAGKVPITGVEGGLERHNSGQSQRSLTLTLPRLSLMNACT 388
Qy 497 EYKCKVSNKALP-----APIEKTS-KAKGQPREPPQVYTLTPPSDELTQKQV 542
Db 389 SVTCTLNHPSLPPORLMALREPAQAQPVLSLNLASSDPPEASW----- 434
Qy 543 SLTCLVKGFPYPSDIAVEMESNGQPENY---YKTTIPVLDSDSGSFYXKLTVDKSRMKG 599
Db 435 -LLCEVSGFSPPNILMLMEDQREVNTSGFAPARPPQPGSTTFAMWSVLVPAPSPPP 493
Qy 600 NVFSGSVNHE---ALHNHYTQKSLSLSPGLQDDETCABAQDELD-----GLMTT 646
Db 494 AYTCTVSHEDSRTLLN--ASRSLEVS-YLMTPLIPGSKDENSDDYTFDVGSLMTT 549

```

## RESULT 93

Q9DBM4

PRELIMINARY; PRT; 130 AA.

```

ID 09DBM4; PRELIMINARY; PRT; 130 AA.
AC 09DBM4;
DT 01-JUN-2001 (TRMBLrel. 17, Created)
DT 01-OCT-2002 (TRMBLrel. 17, Last sequence update)
DE 01-OCT-2003 (TRMBLrel. 25, Last annotation update)
GN 1810027001R1K
OS 1810027001R1K
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; PubMed=11217811;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamata I.,
RA Saico T., Okazaki Y., Gojibori T., Bono H., Kanakawa T., Saico R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Giesl C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsumoto Y., Nishikido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.S., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guelincich S., Hill D., Holtman M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombereite P.,
RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schenbach C., Seta Y., Shihata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wymah-Botie A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK007622; BAB25142.1; -.
DR PIR; S22760; S22760.
DR HSSP; P01842; 7FAB.
DR MGD; MGI:1916426; 1810027001R1K.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IgC1.1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 130 AA; 14253 MW; 438197975E76E54 CRC64;

```

```

Query Match 5.4%; Score 183.5; DB 11; Length 130;
Best Local Similarity 36.5%; Pred. No. 2.7e-06;
Matches 38; Conservative 18; Mismatches 43; Indels 5; Gaps 3;
Qy 520 GQPR-EPQVYTLPPSRDELTKQVSLTCLVKGFPYPSDIAVEMESNGQP-ENNYKTTTPVL 577
Db 25 GQPKSPVTLPPSRSELETNKATLVCTITDFYGVGVTVDMKVDGTVTQGMETTPQSK 84
Qy 578 DSDSGFSLYKLTVDKSRWQGNVPSGSMHEALHNHYTQKSLS 621
Db 85 QSNMKYMASSYLTLTARAMEHRSYSQVTHE---GHTVEKSL 125

```

## RESULT 94

Q8N355

PRELIMINARY; PRT; 234 AA.

```

ID Q8N355; PRELIMINARY; PRT; 234 AA.
AC Q8N355;
DT 01-OCT-2002 (TRMBLrel. 22, Created)
DT 01-OCT-2002 (TRMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TRMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (Apr-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC028090; AAH28090.1; -.
DR PIR; S12441; S12441.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig_2.
DR SMART; SM00409; IG; 2.
DR SMART; SM00407; IG; 1.
DR SMART; SM00406; IG; 1.
DR PROSITE; PS00835; IG_LIKE; 2.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 234 AA; 24792 MW; CC848CAEBA49D63 CRC64;

```

## Query Match

5.4%; Score 183; DB 4; Length 234;

Best Local Similarity 23.9%; Pred. No. 6.8e-06;

Matches 50; Conservative 42; Mismatches 99; Indels 18; Gaps 6;

```

Qy 415 GQPSVFLPPPKOTLMISRTPEVTCVVVDVSHEDPEVKFNWY-----VDGVEVHNACT 468
Db 16 GSVSYVLTQPPSVAVAGTARITCGGNNGSK-----SVHMYQKPGQAFLVYDDSD 71
Qy 469 KPR---EQGNSTYRVSVLTVLVHQMNGEKYCKVSNKALPPI---EXTISKAKGQ 522
Db 72 RPSGIPERFSSNGCNTATTLTISRVDADGADYCCQLWDSSDHPVFGGTLKLTVLQ 131
Qy 523 R-EPQVYTLPPSRDELTKQVSLTCLVKGFPYPSDIAVEMESNGQP-ENNYKTTTPVL 580
Db 132 KAAPSVTLPFPSSSEELQNKATIVCLISDFPGAVTAMKAKDSPPVAKGVETTPSKQSN 191
Qy 581 GSPFLYKLTVDKSRWQGNVPSGSMHE 609
Db 192 NKYAASVSLTPQGMKSHRSYSCQVTHE 220

```

## RESULT 95

Q28260

PRELIMINARY; PRT; 739 AA.

```

ID Q28260;
AC Q28260;
DT 01-NOV-1996 (TRMBLrel. 01, Created)
DT 01-NOV-1996 (TRMBLrel. 01, Last sequence update)

```



DT 01-OCT-2003 (TREMblRel. 25, last annotation update)  
DB Vascular cell adhesion molecule-1.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteheria; Carnivora; Carnivora; Flesipedia; Canis.  
OX NCBI\_Taxid=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ballantyne C.M., Clubb F.J., Perrard J.L., Radencovic B.,  
RA Youker K.A., Smith C.W., Entman M.L., Hawkins H.K., Frazier O.H.,  
RA Willerson J.T.;  
RT "Increased Expression of VCAM-1 and ICAM-1 in Early Cardiac Allograft  
RT Arteriosclerosis in the Dog."  
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.  
DR EMBL; U32086; AA84866.1; -.  
DR HSSP; P19320; 1VCA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0016337; P:cell-cell adhesion; IEA.  
DR InterPro; IPR003987; ICAM VCAM-1.  
DR InterPro; IPR002110; Ig-like.  
DR InterPro; IPR003598; Ig\_G2.  
DR InterPro; IPR003989; VCAM-1.  
DR Pfam; PF00047; Ig\_7.  
DR PRINTS; PRO1472; ICAMVCAM1.  
DR PRINTS; PRO1474; VCAM1.  
DR SMART; SM00408; ICG2; 4.  
DR PROSITE; PS50835; IG\_LIKE; 5.  
KW Immunoglobulin domain.  
SQ SEQUENCE 739 AA; 81412 MW; BB5DA3853739C615 CRC64;

Query Match	5.3%	Score 182.5;	DB 6;	Length 739;
Best Local Similarity	21.5%	Pred. No. 3.8e-05;		
Matches 112;	Conservative 96;	Mismatches 183;	Indels 131;	Gaps 29;

```

Db 237 QEGSSVTMTCSSEGLPPFOIFMSKLDNGNLQL----- 270
Qy 89 QGNPFLIIKNLKIEDSDTYIC---EVEDQEEVOLLV---FGLTANSDTHLQ--GQS 138
Db 271 SGNALTLTLARLEDSGTIVYCEGVANEVKDQKEVELLIVOEKPFYIEISPGOIIAQIDS 330
Qy 139 LTLTLESPPGSSPS-----VOCSPRG--KNIOGGK--TIVSOLTLODSGTMTCTYLOMQK 191
Db 331 VLTGVTDCSPSPSSMTQIDSLPSLGTIVKYGAKSTLTLSPVNLNENHSYLCVTGCHK 390
Qy 192 KYEPKIDIVLAFQK-----ASSIYKKEGEQVEFSFPLAFVEXLXTSGELMJOABRAS 246
Db 391 KLEKGIKVDLVSFPDPREVEMSGLLVNDGNPVTVSCEVENVPDBRL--EIEL-FXGETII 447
Qy 247 SSKSMI-----TPLNKNKEVSVKRVNTOQDPKLOMKKL---PLHLTLPLQALPOYAGSN 296
Db 448 ESKSPLEEDMDKKSLTETKSLMTFIPITTD---TKVLVCLAKHIDIMEFEPPKORGS-- 501
Qy 297 LTLALEATGKXLHQEVNLYVMRATOLQ---NLTCSEVGPSPKYLMLSKLENNEAVVS 352
Db 502 -TQTLVYVAVP--RDTLVVSPSSIVESGSPPNMTCSSDGLPAPNIIIMSRRLSNRLO--S 557
Qy 353 KREKVVWVLN---PEAGMOQL--LSDSGOVLLENIKVLPWMSFVPEKSDCKHTHTCP 407
Db 558 LSEEDILTLTSAKMDSGIYCEGIGNQG--ISRKEVELI---IQVAPKOI----- 603
Qy 408 CPAPELLGSPSYFLPPPKPOTLMSIRPDEVTCVVDVSHEDPEYKFMWYV--DGEVYHN 465
Db 604 ----QLINPBEVSV--KEGDTVIIS-----CTC-----GNVPKXMIILKKAELFGD 643
Qy 466 AKTKPREQYNSTYRVVSVLTVLHQDWL--NGKEYCKVSNKA 506
Db 644 TULKSRLDAY-----TIHVOUEDAGVYCESKNEA 674

```

RESULT 96  
Q63669

ID	063669	PRELIMINARY;	PT;	739 AA.
AC	063669			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DE	Vascular cell adhesion molecule 1 precursor.			
	VCAM-1.			
OS	<i>Rattus norvegicus</i> (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_Taxid=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Sprague Dawley; TISSUE=Lung;			
RX	MEDLINE=92305064; Pubmed=1377031;			
RA	Williams A.J., Atkins R.C., Fries J.W., Gahbrone M.A.Jr.,			
RA	Cybulsky M.I., Collins T.;			
RT	"Nucleotide sequence of rat vascular cell adhesion molecule-1 CDNA."			
RL	Biochim. Biophys. Acta 1131:214-216(1992).			
DR	EMBL; X63722; CAA45254.1; -.			
DR	HSSP; P19320; IVCA.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0016337; P:cell-cell adhesion; IEA.			
DR	InterPro; IPR003987; ICAM VCAM-1.			
DR	InterPro; IPR007110; IG-like.			
DR	InterPro; IPR003599; IG c2.			
DR	InterPro; IPR003989; VCAM-1.			
DR	Pfam; PF00047; Ig; 5.			
DR	PRINTS; PRO1472; ICAMVCAM1.			
DR	PRINTS; PRO1474; VCAM1.			
DR	SMART; SMO0408; IGC2; 4.			
DR	PROSITE; PS50835; IG LIKE; 5.			
KW	Immunoglobulin domain; Signal			
FT	SIGNAL			
	1	24		
FT	POTENTIAL.			
FT	CHAIN	25	739	
SEQUENCE	739 AA;	81136 MW;	ASAAD1172F67FB96 CRC64;	
	VASCULAR CELL ADHESION MOLECULE 1.			

Query Match	5.3%;	Score 182.5;	DB 11;	Length 739;
Best Local Similarity	20.8%;	Pred. No. 3.8e-05;		
Matches 124; Conservative	87;	Mismatches 195;	Indels 191;	Gaps 28;

```

QY      33 KGDVTELTCTASQOKKSIQFHMKJNSNQIKILGNQSGFLTKPGSKCNDAADSRSLMDQGNF 92
Db      238 EGAATYMTQASBGLPAPETWFSKK-----LDNGVQLL-----SGNA 274
QY      93 PLIKNKLEIEDSDPYICE---VEDOKREVOVLV-----FGLTANBDTHLLQOSLTLT 142
Db      275 TLTLIARMEBDSGIYCEGVNLVERDKTEVELIYQEKFPYVIDBSGSAVAQVDSVLT 334
QY      143 LESPPGSSPSYQCSPPKGNIOG-----GKITSVSOLELODSGTWTCTYLOKQYEF 195
Db      335 CAAVGCDSPSPSMRTOTDSPLVNGVRDEGATSTLTLPVGEDHSYLCVTYTCORRLEK 394
QY      196 KIDIYVLAFOASSIYVKEGEQVEFSPPLAFVTEKLTGSGELMWQABRASSKSMITFD 255
Db      335 TIQVEYVSF-----PEDPELEISGPLVH----- 417
QY      256 LKNKEVSVKRVTDPKLOMGKKLPLHLTLPOALFOYAGSGLTLALBAKTGKLHQEVNLY 315
Db      418 -----GRPVYTNCTVPVNYFPD-----HLEIELL 441
QY      316 VMRATOLOKULTCVEWGPSTPKMLSLKLENKEAKVSKREKPPVNLNPEAGMOCCLSDS 375
Db      442 KGETTLNKKFLREEL-GTKS-----LEKYSLEMT-----FIPTA-----EDT 477
QY      376 GOVLLESNIKULPTWSPFVPEBKSCDKHTTCCPPCAPELLGSPSVFLFP-PRKQDTLMISR 434
Db      478 GKALV--CLATLHSSQMESEPKQOSQOTLYLVNNAFK--EPLTWSSPVPPEGSPVYN 531
QY      435 TPBYTGVVVDVSHEDPEVKFNMYVDQVEVHNAKTKPREBOYNSYTRVVSULTVLHQMNLN 494
Db      532 ---LTC-----SSDGFPPFKILM---SHQLKNGELOQPSQ--NTT---LSFNATWED--- 573

```



```

Qy 495 GKEYCKKSNKALPAPIEKTISK-----AKGQPERPVYTLPPSRDELTKNOVSLTCLV 548
Db 574 SGIVYCEGINEA-----GISKSVELLIGSSMDIQ-TVPFSKSVBEGDTVIISCTC 625
Qy 549 KKFYPSDIAVWESNGQPPNNYKTTTPVLDSDGSEFLYSKLTVDKSRWQGNVFCSC 604
Db 626 -GSVEIWI-----LKKAKTGDMVLKSYNGSY-----TIKKAQLQADGVYEC 668

RESULT 97
Q95YM2 PRELIMINARY; PRT; 17352 AA.
ID Q95YM2;
AC Q95YM2;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE I-connectin.
GN I-CON.
OS Procamburus clarkii (Red swamp crayfish).
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacidea; Cambaridae; Procamburus.
OC NCBI_TaxId=6728;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423462; PubMed=11532946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RA Sun P., Maruyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RT sarcomeres of crayfish claw muscle.";
RL EMBO J. 20:4826-4835(2001).
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: AB055861; BAB64297.1; -.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro: IPR000577; FGGY_kin.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR007110; IG-II-like.
DR InterPro: IPR003598; IG_c2.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00047; Ig; 49.
DR Pfam: PF00018; SH3; 1.
DR SMART: SM00408; IGC2; 13.
DR SMART: SM00325; SH3; 1.
DR PROSITE: PS00933; FGGY_KINASES_1; 3.
DR PROSITE: PS50835; IG_LIKE; 49.
DR PROSITE: PS50002; SH3; 1.
DR Immunoglobulin domain; SH3 domain.
SQ SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;

Query Match 5.3%; Score 182; DB 5; Length 17352;
Best Local Similarity 19.9%; Pred. No. 0.0038;
Matches 128; Conservative 103; Mismatches 227; Indels 184; Gaps 29;

Qy 37 VELTCTASGKSIQPHMKSNQIKLGNQSLTGKPSKLNRAASRLMVGQNPFLII 96
Db 16247 VEVV---NEKATVTVM-KDGEIITEHHEFPKLVSGQKTR-----KVLV 16285

Qy 97 KNLKIEDSDPTVCEVEDQKEEVQLVFGTLTAN-----SDTHLLQGSLLTLLESPGSS- 150
Db 16286 MEATISDEBEYTCVLDGDCETALVELYRELPAEIVKMKDQVVSXKDRATMEVELTKGDV 16345

Qy 151 -----PSVOCR-----SPRKNIGQKTLISVSQLELDGSGTWCTVQLQNKVVEFKID 198
Db 16346 ITWYKDEYVIRFSDHYQLSIDK---VQRLWYVNCQFEDSGTYRAVVGKSECSATLKVE 16401

Qy 199 IYVLAFOKASSTIVYKKEG-----QVRSFPLATVEKLTGSGELW---QAEKAS 246
Db 16402 LQV-----EGDPSKLLPAQMDVNFKTDATFVE--ITQDYEVKWLREGAELSS 16447

Qy 247 SSKSWITFDLKNKEVSVKRVTDOPKQLQ-----MGKKLPLHLTLPOALPYAGSGNLT 299

```

```

Db 16448 SEKYVKEAKKRILIVKSVQTDACEYSVGLNLTSCVLHVVRMETAP----- 16497
Qy 300 ALEAKTGLHDEVNLVWRBARQLOKNLTCEV--WGPTSPKLM-----LSLLENKE---A 349
Db 16498 ---KTFPEHOKVEYLVYKG---KDAVLKLPFTATPPKFWYHKGOLLTENTEKLLP 16549
Qy 350 KYSKREKEV---WVLNPEAGMOC-LTSDSGVLLSNIKYLPMTWSTPEVPSKCDKTHTC 405
Db 16550 TISQEBASITIKQVENIDCGEYRLKLCNDCCAAVADFTLKLDKPSQGTCEPMEVNTV- 16608
Qy 406 PCCPAPELLGGPSVFLPPPKKDTLMTSRTEVTCVVVDVSHEDPEVKFNMYVDGVEVHN 465
Db 16609 -----SVTLHMSLPKE---DGRAITNYIVEVSKKEQ---SWVFNOQV-- 16647
Qy 466 AKTKPREQVNSTYRVSVVLVLDQMDLNGEYCKKSNKALPAPIEKTISKAGQPREP 525
Db 16648 -----KITQSTTVVTSLT-HH-----EYCFRVS-----AENETIRSEKSNASR 16695

Qy 526 QVYTLPPSRDELTKNOVSLTCLVKGFPYPSDI-----AVWESNGQPEN----- 568
Db 16686 YVKCEPRARPPVVKEQLOAVTGLHGEVVLRCVVTATPTPKIDMLKDGRFVSGNTSYE 16745
Qy 569 NYKTTTPVLDSDGSEFLYSKLTVDKSRWQGNVFCSCVWMEA 610
Db 16746 NFTA-----LTIKETMETSGMYTCRASNBA 16772

RESULT 98
Q8ISF3 PRELIMINARY; PRT; 2693 AA.
ID Q8ISF3;
AC Q8ISF3;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE 301kDa_2 protein.
GN ISOF.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; PubMed=12381307;
RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Domains, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions.";
RL J. Mol. Biol. 323:533-549(2002).
DR EMBL: AY130758; AAN61521.1; -.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO: GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO: GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR007110; IG-II-like.
DR InterPro: IPR003598; IG_c2.
DR InterPro: IPR000719; Prot_kinase.
DR InterPro: IPR02290; Ser_thr_pkinase.
DR InterPro: IPR008271; Ser_thr_pkin_AS.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00047; Ig; 19.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Prot_kinase; 1.
DR SMART: SM00060; FN3; 1.
DR SMART: SM00408; IGC2; 19.
DR SMART: SM00220; S_TKC; 1.
DR SMART: SM00219; TYKC; 1.
DR PROSITE: PS50835; IG_LIKE; 18.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.

```



SQ SEQUENCE 2693 AA; 299094 MW; 1CE4AD2011E8391A CRC64;  
 Query Match 5.3%; Score 180.5; DB 5; Length 2693;  
 Best Local Similarity 20.2%; Pred. No. 0.00034;  
 Matches 137; Conservative 99; Mismatches 254; Indels 189; Gaps 33;

QY 28 VLKGGKDTVELTCTASQKSIQPHWKNNSQIKILGNQSGFLTKGPSKLNDRADSRSLW 87  
 DB 1100 VLKTAGETATFTCGSYANPAQVWV-----LHNGKALQQTNSYKTRLF 1144  
 QY 88 DQGNPLIKKLKIEDSTYICEVEDQKEV-----QLLVGLTANS----- 129  
 DB 1145 DDMTATLIVENVDELGGTYTAVANNQPGDVHTSAQLTISGEAKKIAASLPYITIELKP 1204  
 QY 130 DTHLLQGSGLTLTLSPGSSP-----SVQCRSP-----GKNIQSGKTLVSQ 173  
 DB 1205 KINVEG--ATLSIQADLNGSPPEVWVLKNSLVEDRLOMKCDGNYQ-----LVVRD 1258  
 QY 174 LELQDSGTWCTVVLQNGKVEFKIDIVLAFQKASSIVYKKEGQVEF-----S 222  
 DB 1259 VGLDEGTYYITTAENEKGIKQNTVEVSV---TKSEYKEKKEKKVEKDEGKKKPGRPG 1315  
 QY 223 FPL---AFTVEKLT-----GSGELMWAERASSKSWITF-DLNKEVSVK----- 264  
 DB 1316 LPRPSGASKTEQVTWAFDAPSEGPADSYEVERRCPDQREWVSCGSTLSLELEIKGLTPNT 1375  
 QY 265 ----RVTDPKLQMGKLLPLHLTLPOA----LPQYAGSGNLTALAEAKTKLHOENVLV 316  
 DB 1376 EYIRVAVGKNGKQGLGESEMTSTLKTASVGAPOF-----TISPSQK-----IIA 1420  
 QY 317 MRATQLOKNTLCEVWGTPSPKMLSLKLENKAKVSKREK-----PWWVLN---PEA 365  
 DB 1421 NRDEFE--IAVFSGTPPTP-----SVKWKENLQIVPEKIDIVATTSTSSILNKSQEN 1474  
 QY 366 GMMQCLL-SDSGOVLLESNIKVLPTWSTPVEPKSCDKT--HTCPCPAPEL----- 413  
 DB 1475 GTFENCLILENLGQASASCQVTIFN-----KPSLQSTPHSLERNLVPFLQKALNNEA 1528  
 QY 414 LGSPSVFLFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVNAKTKPRE 473  
 DB 1529 QAGQIWL-----TCRISRSSESTVA-----WPKDDRISAGRYELSS 1567  
 QY 474 QYNSTYRVSVLVTLHODMLNGEKYCKVSNKALPAPIEKTISAKAQPREPQVYTLPPS 533  
 DB 1568 DKSNHKLIV-CHAVQSD--TGG-YRCVVTNKYGAESECNVAV-----EDVTKFIAPS 1617  
 QY 534 -----RDELTKNQVSLTCLVKGFPSPDIADVEMESNGQPENNYKTPVLDSDGSPFLY 586  
 DB 1618 FSATLSDSTALIGHNITLCEKVEGSPADE--VSWTKQSERISTTRIRIQTQDENG--- 1671  
 QY 587 SKLTVDKSRMOQGNVFS 605  
 DB 1672 CKLSTSKAESDDMGVYCS 1690

RESULT 99  
 Q8ISF4 PRELIMINARY; PRT; 2708 AA.  
 AC Q8ISF4:  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE 301kDa\_1 protein.  
 GN ISOF.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22269627; PubMed=12381307;  
 RA Flaherty D., Gernert K., Simeleva N., Tang X., Mercer K.,  
 Borodovsky M., Benian G.;

RT "tilins in Caenorhabditis elegans with Unusual Features: Coiled-coil  
 RT Domains, Novel Regulation of Kinase Activity and Two New Possible  
 RT Elastic Regions.";  
 RL J. Mol. Biol. 323:533-549(2002).  
 DR EMBL: AY130758; AAN61520.1; -.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0004674; F:Protein serine/threonine kinase activity; IEA.  
 DR GO: GO:0004713; F:Protein-tyrosine kinase activity; IEA.  
 DR GO: GO:0006468; P:Protein amino acid phosphorylation; IEA.  
 DR InterPro: IPR003961; FN III.  
 DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR003598; Ig\_c2.  
 DR InterPro: IPR000719; Prot\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR008271; Ser\_thr\_kinase.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00041; fn3.2.  
 DR Pfam: PF00047; Ig; 19.  
 DR Pfam: PF00069; pkinase; 1.  
 DR ProDom: PD000001; Prot\_kinase; 1.  
 DR SMART: SM00060; FN3; 1.  
 DR SMART: SM00408; IGG2; 19.  
 DR SMART: SM00220; S\_TKC; 1.  
 DR SMART: SM00219; TYRC; 1.  
 DR PROSITE: PS50835; IG\_LIKE; 18.  
 DR PROSITE: PS50107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 SQ SEQUENCE 2708 AA; 300729 MW; 2B8EB8656FA571 CRC64;

Query Match 5.3%; Score 180.5; DB 5; Length 2708;  
 Best Local Similarity 20.2%; Pred. No. 0.00035;  
 Matches 137; Conservative 99; Mismatches 254; Indels 189; Gaps 33;

QY 28 VLKGGKDTVELTCTASQKSIQPHWKNNSQIKILGNQSGFLTKGPSKLNDRADSRSLW 87  
 DB 1115 VLKTAGETATFTCGSYANPAQVWV-----LHNGKALQQTNSYKTRLF 1159  
 QY 88 DQGNPLIKKLKIEDSTYICEVEDQKEV-----QLLVGLTANS----- 129  
 DB 1145 DDMTATLIVENVDELGGTYTAVANNQPGDVHTSAQLTISGEAKKIAASLPYITIELKP 1219  
 QY 130 DTHLLQGSGLTLTLSPGSSP-----SVQCRSP-----GKNIQSGKTLVSQ 173  
 DB 1220 KINVEG--ATLSIQADLNGSPPEVWVLKNSLVEDRLOMKCDGNYQ-----LVVRD 1273  
 QY 174 LELQDSGTWCTVVLQNGKVEFKIDIVLAFQKASSIVYKKEGQVEF-----S 222  
 DB 1274 VGLDEGTYYITTAENEKGIKQNTVEVSV---TKSEYKEKKEKKVEKDEGKKKPGRPG 1330  
 QY 223 FPL---AFTVEKLT-----GSGELMWAERASSKSWITF-DLNKEVSVK----- 264  
 DB 1331 LPRPSGASKTEQVTWAFDAPSEGPADSYEVERRCPDQREWVSCGSTLSLELEIKGLTPNT 1390  
 QY 265 ----RVTDPKLQMGKLLPLHLTLPOA----LPQYAGSGNLTALAEAKTKLHOENVLV 316  
 DB 1391 EYIRVAVGKNGKQGLGESEMTSTLKTASVGAPOF-----TISPSQK-----IIA 1435  
 QY 317 MRATQLOKNTLCEVWGTPSPKMLSLKLENKAKVSKREK-----PWWVLN---PEA 365  
 DB 1436 NRDEFE--IAVFSGTPPTP-----SVKWKENLQIVPEKIDIVATTSTSSILNKSQEN 1489  
 QY 1436 NRDEFE--IAVFSGTPPTP-----SVKWKENLQIVPEKIDIVATTSTSSILNKSQEN 1489  
 QY 366 GMMQCLL-SDSGOVLLESNIKVLPTWSTPVEPKSCDKT--HTCPCPAPEL----- 413  
 DB 1490 GTFENCLILENLGQASASCQVTIFN-----KPSLQSTPHSLERNLVPFLQKALNNEA 1543  
 QY 414 LGSPSVFLFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVNAKTKPRE 473  
 DB 1544 QAGQIWL-----TCRISRSSESTVA-----WPKDDRISAGRYELSS 1582  
 QY 474 QYNSTYRVSVLVTLHODMLNGEKYCKVSNKALPAPIEKTISAKAQPREPQVYTLPPS 533  
 DB 1583 DKSNHKLIV-CHAVQSD--TGG-YRCVVTNKYGAESECNVAV-----EDVTKFIAPS 1632



```
QY 534 -----RDELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTPPVLDSDGSFPLY 586
DB 1633 FSATLSDSTALLGHNITLLECKVEGSPAPE--VSWTKDGERISTTRIRIQTQDENG----- 1686
QY 587 SKLTVDKSRMQGQNVFSCS 605
DB 1687 CKLSISKAESDDMGVYVCS 1705

RESULT 100
Q8MNS0 PRELIMINARY; PRT; 2780 AA.
AC Q8MNS0;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
GN F12F3.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Br1stol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Br1stol N2;
RA Fulton B., Wohlmann P.;
RT "The sequence of C. elegans cosmid F12F3.";
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Br1stol N2;
RA Waterston R.;
RT Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
RL EMBL; U80022; AAA29672.1; -.
DR WormRep; F12F3.2a; CE30753.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004674; P:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; P:protein-tyrosine kinase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR003962; FNIII subd.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; IG-1-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002390; Ser_thr_kinase.
DR InterPro; IPR008271; Ser_thr_kin_AS.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; ig; 19.
DR Pfam; PF00069; pkinaee; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR PRODOM; PD000001; Prot_kinase; 1.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGC2; 19.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYKc; 1.
DR PROSITE; PS50835; IG_LIKE; 18.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW Hypothetical protein; ATP-binding; Immunoglobulin domain; Repeat;
KW Transferase.
```

```
SEQ SEQUENCE 2780 AA; 308646 MW; 046D057107B935C1 CRC64;
Query Match 5.3%; Score 180.5; DB 5; Length 2780;
Best Local Similarity 20.2%; Pred. No. 0.00036;
Matches 137; Conservative 99; Mismatches 254; Indels 189; Gaps 33;

QY 28 VLGGKGPVVELTCTASOKKSIOFMKNSNQIKILGQSGFLTXGPKLNDRADSRSLW 87
DB 1187 VLKTAGEETATFTQSYANPAQYVW-----LHNGKLLQGTQSKYKTRLF 1231
QY 88 DQGNPLIIKNKLIKEDSDTYICEVEDQKEV---QLLVGLTNS----- 129
DB 1232 DQNTATLVIEENVDELQGTVAVANNOFGDVHTSAQLTSGSEAKKIAASLPYFIELKP 1291
QY 130 DTHLLQGSLLTLTLESPPGSSP-----SYQCSPP-----GKNLQSGTLTYSQ 173
DB 1292 KINVEG--ATLSTIQADLNGSPIREVWLKNSLVSBDRIQMKCDGNYO---LTVRD 1345
QY 174 LELODSGTWCTCTVQNGKVEFKIDIVLAFQKASSIYKKKGQVEF-----S 222
DB 1346 VGLDEGTYITTAENKGIQNTVEGV---TKSEYKEKKEKKVEKDGKKKPKGRPG 1402
QY 223 FPL--AFTVEKLT-----GSGELMWQAEKASSSKSWITF-DLKNKEVVK----- 264
DB 1403 LPRPSGASKTEQVTMAFDAPSEGPADSYEVERRCPDQREWVSCSTKSLLEIKLTPNT 1462
QY 265 ----RYQDPRLQMGKPLHLTLPPQ---LPQYASGNLTLLAEKTKGLHQEVNLV 316
DB 1463 EYIFRVAGKNQGGISEMSTTKTASVGAQPF-----TISPQK-----IIA 1507
QY 317 MRATOLQNLTCFVWGPTSPPLMLSLKLENKAVSKREK-----PWLVN---PEA 365
DB 1508 NRDEFE--IATVPSGTPF-----SVKMYKENLQIVDEKIDVATTSSTLSLNKSOEN 1561
QY 366 GMMOCLL-SDSGQVLLSINIKVLPWTSTPYEPKSCDKT--HTCPPAPAEI----- 413
DB 1562 GFNCLLENLGGQASACQVTFIN-----KPAISLQSTPDHSLBRNLVPTLQKALNNEA 1615
QY 414 LGSPSVLPFPKPDITLMIISTPEVTCVVVDVSHEDPEVKKNVYVDGEVNAATKPRE 473
DB 1616 QAGQQIML-----TRISSRSSTVA-----WPKDERISARRELSS 1654
QY 474 QYNSTYRVSVLTVLHODWLNGKEYKCKVSKALPAPIEKTISKAKGPRPQVYTLPPS 533
DB 1655 DKSNHKLV-CHAVSQSD--TGR-YRCVYTKKGYASECVAV-----EDYKFIAP 1704
QY 534 -----RDELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTPPVLDSDGSFPLY 586
DB 1705 FSATLSDSTALLGHNITLLECKVEGSPAPE--VSWTKDGERISTTRIRIQTQDENG----- 1758
QY 587 SKLTVDKSRMQGQNVFSCS 605
DB 1759 CKLSISKAESDDMGVYVCS 1777
```

Search completed: August 3, 2004, 13:13:21  
Job time : 55.4955 secs



**THIS PAGE BLANK (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:09 ; Search time 65.926 Seconds  
(without alignments)  
2777.216 Million cell updates/sec

Title: SEQ3  
Perfect score: 3414  
Sequence: 1 NMRGVFRLHLVLAQLALP.....DETCALAQNGDLGWLTTDP 648

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 125 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseq1980s:\*  
2: geneseq1990s:\*  
3: geneseq2000s:\*  
4: geneseq2001s:\*  
5: geneseq2002s:\*  
6: geneseq2003as:\*  
7: geneseq2003bs:\*  
8: geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3268.5	95.7	631	1 AAP93009	Aap93009 Genetic C
2	3268.5	95.7	631	1 AAB19508	Aab19508 CD4-IGG1
3	3268.5	95.7	631	3 AAY51079	Aay51079 Human fus
4	3268.5	95.7	631	3 AAY59169	Aay59169 CD4-Ig fu
5	3219.5	94.3	729	1 AAP93008	Aap93008 Genetic C
6	3219.5	94.3	729	3 AAB19507	Aab19507 CD4-IGG1
7	3219.5	94.3	729	3 AAY59168	Aay59168 CD4-Ig fu
8	3209.5	94.0	729	3 AAY51078	Aay51078 Human fus
9	3169	92.8	616	3 AAY51082	Aay51082 Human fus
10	3169	92.8	616	3 AAY59172	Aay59172 CD4-Ig fu
11	3168	91.9	614	1 AAP93012	Aap93012 Genetic C
12	3077	90.1	616	3 AAB19511	Aab19511 CD4-IGG1
13	2268	66.4	534	2 AAR26531	Aar26531 Sequence
14	2268	66.4	534	2 AAR26530	Aar26530 Sequence
15	2156.5	63.12	450	6 AAB37574	Aab37574 Human TID
16	2151	63.0	530	2 AAR26763	Aar26763 Human TID
17	2151	63.0	530	2 AAY85080	Aay85080 CD4-IGG2
18	2151	63.0	530	4 AAB67323	Aab67323 CD4-IGG2
19	2151	63.0	530	4 AAB80884	Aab80884 Human CD4
20	2151	63.0	530	6 AABG7123	Aabg7123 CD4-Igmu
21	2141	62.7	530	2 AAR46679	Aar46679 CD4-IGG2
22	2137	62.3	449	6 AAB37576	Aab37576 Human TID
23	2077	60.8	432	2 AAR26762	Aar26762 CD4-gamma
24	2077	60.8	432	2 AAR46678	Aar46678 CD4-gamma
25	2077	60.8	432	3 AAY85079	Aay85079 Human/CD4

26	2077	60.8	432	4 AAB67322	Aab67322 CD4-gamma
27	2077	60.8	432	4 AAB80883	Aab80883 Human CD4
28	2077	60.8	432	6 AABG7122	Aabg7122 CD4-gamma
29	2066	60.5	481	3 AAB19510	Aab19510 CD4-IgM f
30	2066	60.5	481	3 AAY59171	Aay59171 CD4-Ig fu
31	2060	60.3	481	3 AAY51081	Aay51081 Human fus
32	2058	60.3	481	1 AAP93011	Aap93011 Genetic C
33	2053.5	60.1	436	3 AAY51080	Aay51080 Human fus
34	2047.5	60.0	474	3 AAY59170	Aay59170 CD4-Ig fu
35	2039	59.7	532	2 AAR27278	Aar27278 CD4-gamma
36	2039	59.7	532	2 AAR78678	Aar78678 T-cell re
37	2039	59.7	532	2 AAR89458	Aar89458 CD4-eta f
38	2039	59.7	532	2 AAW83141	Aaw83141 Chimeric
39	2039	59.7	575	2 AAR27276	Aar27276 CD4-zeta
40	2039	59.7	575	2 AAR78676	Aar78676 T-cell re
41	2039	59.7	575	2 AAR89456	Aar89456 CD4-zeta
42	2039	59.7	575	2 AAW02213	Aaw02213 CD4-T-cel
43	2039	59.7	575	2 AAW83140	Aaw83140 Chimeric
44	2038	59.7	458	3 AAB07769	Aab07769 DNA encod
45	2035	59.6	462	2 AAR27277	Aar27277 CD4-eta p
46	2035	59.6	462	2 AAR78677	Aar78677 T-cell re
47	2035	59.6	462	2 AAR89457	Aar89457 CD4-gamma
48	2035	59.6	462	2 AAW02214	Aaw02214 CD4-FC re
49	2035	59.6	462	2 AAW83142	Aaw83142 Chimeric
50	2035	59.6	532	2 AAW02215	Aaw02215 CD4-T-cel
51	2030	59.5	400	2 AAR06374	Aar06374 Truncated
52	2030	59.5	458	1 AAR01390	Aar01390 Clone pT4
53	2030	59.5	458	1 AAP91369	Aap91369 T4 protei
54	2030	59.5	458	2 AAY39826	Aay39826 Soluble h
55	2030	59.5	2037	2 AAR04032	Aar04032 Full leng
56	2030	59.5	2037	2 AAR07641	Aar07641 Full leng
57	2029	59.4	350	2 AAR89450	Aar89450 CD4 D1-D4
58	2026	59.3	398	2 AAR78673	Aar78673 CD4 domai
59	2024	59.3	416	3 AAB19509	Aab19509 CD4-IgM f
60	2024	59.3	458	3 AAY88329	Aay88329 T4 glycop
61	2024	59.3	458	4 AAB81502	Aab81502 Human CD4
62	2024	59.3	458	7 ADD25609	Add25609 Binding d
63	2024	59.3	458	7 ADE57489	Ad57489 Human Pro
64	2024	59.3	473	7 ADA44807	Ada44807 CD4/TCR C
65	2024	59.3	519	7 AAR20152	Aar20152 Human CD4
66	2022	59.2	458	2 AAR13491	Aar13491 Human CD4
67	2018	59.1	394	1 AAP93506	Aap93506 Derived s
68	2017	59.1	402	1 AAP91922	Aap91922 Sequence
69	2017	59.1	402	1 AAP94757	Aap94757 Sequence
70	2016	59.1	458	4 AAG79087	Aag79087 Amino aci
71	2015	59.0	394	3 AAY88328	Aay88328 T4 glycop
72	2012	58.9	394	2 AAY39825	Aay39825 Soluble h
73	2006	58.8	458	7 ADE65841	Ad65841 Human CD4
74	2002.5	58.7	2458	2 AAR07640	Aar07640 Deduced p
75	2002.5	58.7	2458	2 AAR04031	Aar04031 Full leng
76	2002	58.6	458	2 AAR06373	Aar06373 T4 encod
77	2002	58.6	524	1 AAP94703	Aap94703 Sequence
78	2001	58.6	394	3 AAB07768	Aab07768 The solub
79	2000	58.6	399	2 AAR20151	Aar20151 Chimpane
80	2000	58.6	458	2 AAR11285	Aar11285 gp120 bin
81	2000	58.6	458	2 AAR10988	Aar10988 Chimpane
82	1995	58.4	400	2 AAR20150	Aar20150 Chimpane
83	1994	58.4	458	2 AAR04910	Aar04910 T4 protei
84	1982	58.1	399	1 AAP93010	Aap93010 Genetic C
85	1955	57.3	400	2 AAR20154	Aar20154 SOL rhes
86	1930	56.5	458	2 AAR15149	Aar15149 CD4 coord
87	1928.5	56.5	729	2 AAR41042	Aar41042 CD4-GBPH
88	1917	56.2	384	1 AAP90833	Aap90833 Amino aci
89	1915	56.1	942	1 AAR41041	Aar41041 CD4-GBP13
90	1911	56.0	1786	2 AAR41043	Aar41043 CD4-EBA17
91	1904	55.8	434	1 AAP96151	Aap96151 Sequence
92	1904	55.8	434	1 AAP93557	Aap93557 Fusion of
93	1903	55.7	375	2 AAR07721	Aar07721 Recombina
94	1901	55.7	433	2 AAW41376	Aaw41376 Human CD4
95	1896	55.5	433	3 AAY54500	Aay54500 Amino aci
96	1893	55.4	432	2 AAR74422	Aar74422 Eptlope o
97	1893	55.4	432	2 AAY30514	Aay30514 Predicted
98	1893	55.4	432	6 ADA25188	Ada25188 CD4 eptlo



99	1891	55.4	369	2	AAV39824
100	1891	55.4	369	3	AAV88327
101	1891	55.4	370	1	AAV93528
102	1891	55.4	370	4	AAV83356
103	1879	55.0	370	2	AAV12956
104	1878	55.0	370	2	AAV12954
105	1878	55.0	370	2	AAV12964
106	1878	55.0	370	2	AAV12965
107	1877	55.0	370	2	AAV12958
108	1874	54.9	370	2	AAV12951
109	1872	54.8	370	2	AAV12957
110	1871	54.8	370	2	AAV12959
111	1871	54.8	370	2	AAV12962
112	1870	54.8	435	1	AAV90992
113	1866	54.7	370	2	AAV12966
114	1865	54.6	370	2	AAV12952
115	1864	54.6	370	2	AAV12967
116	1863	54.6	435	1	AAV91289
117	1862	54.5	370	2	AAV12955
118	1861	54.5	370	2	AAV12963
119	1861	54.5	370	2	AAV12960
120	1859	54.5	433	2	AAV08335
121	1858	54.4	370	2	AAV12961
122	1858	54.4	400	2	AAV20149
123	1857	54.4	458	2	AAV10987
124	1856	54.4	458	2	AAV20148
125	1852	54.2	370	2	AAV12953

## ALIGNMENTS

## RESULT 1

ID AAV93009 standard; protein; 631 AA.

XX AAV93009;

DT 25-MAR-2003 (revised)

DT 02-NOV-1992 (first entry)

DE Genetic construct which encodes CD4 linked to human IgG1 at the Esp site upstream of the hinge region (fusion protein CD4E-gamma-1).

KW Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy.

KW diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.

XX Homo sapiens.

XX EP325262-A.

XX 26-JUL-1989.

XX 20-JAN-1989; 89EP-00100913.

XX 22-JAN-1988; 88US-00147351.

XX (GENO ) GEN HOSPITAL CORP.

XX Seed B;

XX WPI; 1989-214472/30.

XX N-PSDB; AAN90357.

PT Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV infections or detecting HIV or SIV in sample.

PS Example; Table 2, Page 24-33; 68pp; English.

CC The fusion protein genes of the invention comprise cDNA sequences which encode CD4 or a fragment which binds gp120 ligated to an expression plasmid which encodes an antibody in which the variable region of the gene has been deleted (see WO87-02671). The CD4 portion of the fusion

CC protein may comprise the complete CD4 sequence, the 370 AA extracellular region and the membrane spanning domain, or the extracellular region. The CC Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are CC specifically claimed: fusion proteins CD4E-gamma-1, CD4Mmu, CD4E-gamma-1, and CD4Mmu (No. 67608), pCD4E-gamma (No. 67609) and pCD4E-gamma-1 (No. 67610). The plasmid containing (pCD4E-gamma-1) has been deposited in CC E. coli (MC1061/P3) at the ATCC under accession number 67610. (Updated on CC 25-MAR-2003 to correct PA field.)

XX Sequence 631 AA;

SO Query Match 95.7%; Score 3268.5; DB 1; Length 631;  
Best Local Similarity 99.0%; Pred. No. 1.8e-170;  
Matches 624; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY	1	MNRGVPFRHLIVLQALPAATQGNVVLGKGDVLTCTASOKSIOFHMNSQIK	60
DB	1	MNRGVPFRHLIVLQALPAATQGNVVLGKGDVLTCTASOKSIOFHMNSQIK	60
QY	61	ILNQGSLFKGSPKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEVQL	120
DB	61	ILNQGSLFKGSPKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEVQL	120
QY	121	LVFGLTANSDTHLLQGGSLTLLTLESPGSSPSVQCRSPRGKNIQSGKTLVSQLELDG	180
DB	121	LVFGLTANSDTHLLQGGSLTLLTLESPGSSPSVQCRSPRGKNIQSGKTLVSQLELDG	180
QY	181	TWTCVIVQNKQKVEFKDIIVLAFQKASSTVYKKEGQVSPFLATVETLTSGGLMW	240
DB	181	TWTCVIVQNKQKVEFKDIIVLAFQKASSTVYKKEGQVSPFLATVETLTSGGLMW	240
QY	241	QABRASSKSMITFDLKNKEVSRYVQDPKLCQKGLPLHLTPQALPOYAGSGNTLA	300
DB	241	QABRASSKSMITFDLKNKEVSRYVQDPKLCQKGLPLHLTPQALPOYAGSGNTLA	300
QY	301	LEAKTGKHOEVLVVRATQLOKNTLCBYWGPTSPKMLSLKLENKARVSKREKVMV	360
DB	301	LEAKTGKHOEVLVVRATQLOKNTLCBYWGPTSPKMLSLKLENKARVSKREKVMV	360
QY	361	LNPAAGWQCLLSGQVLLSNTKVLPTWSTPVHADPEBPKSCDKTHTCPCPAPBLG	420
DB	361	LNPAAGWQCLLSGQVLLSNTKVLPTWSTPVHADPEBPKSCDKTHTCPCPAPBLG	420
QY	416	GPSVFLPPPKKOTLMISRTPEVTCVVVDVSHEDPEVKFMVYDGVENHNAKTRPREQY	475
DB	421	GPSVFLPPPKKOTLMISRTPEVTCVVVDVSHEDPEVKFMVYDGVENHNAKTRPREQY	480
QY	476	NSTYRVVSVLTVLHQMNLNGEKYCKVSNKALPAPIEKTIISKAKGPREPOVYTLPSRD	535
DB	481	NSTYRVVSVLTVLHQMNLNGEKYCKVSNKALPAPIEKTIISKAKGPREPOVYTLPSRD	540
QY	536	ELTKNOVSLTCVYKGFPSDIAVWESNGQPENNYKTTPVLDSDGSFPLYSKLTIVDKSR	595
DB	541	ELTKNOVSLTCVYKGFPSDIAVWESNGQPENNYKTTPVLDSDGSFPLYSKLTIVDKSR	600
QY	596	WQGNVPSCSVMHEALNHYTQKSLSLSPG	625
DB	601	WQGNVPSCSVMHEALNHYTQKSLSLSPG	630

## RESULT 2

ID AAB19508 standard; protein; 631 AA.

XX AAB19508;

DT 09-JAN-2001 (first entry)

DE CD4-IgG1 fusion protein CH4Bgammal.

CC CD4; IgG1; human; CD4Bgammal; fusion protein; immunoglobulin; HIV; SIV; gp120; therapy; diagnosis.



```

OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT Protein 1..395 "CD4 extracellular region"
XX FT Protein /note="CD4 extracellular region"
XX FT Protein 400..631
XX FT Protein /note="IgG1 heavy chain"
XX
XX US6117656-A.
XX
XX 12-SEP-2000.
XX
XX 07-JUN-1995; 95US-00479353.
XX
XX 22-JAN-1988; 88US-00147351.
XX 23-JAN-1989; 89US-00299596.
XX 09-JUN-1992; 92US-00896781.
XX 12-APR-1993; 93US-00057952.
XX 04-FEB-1994; 94US-00191708.
XX
XX (GENO ) GEN HOSPITAL CORP.
XX
XX Seed B;
XX
XX WPI; 2000-586558/55.
XX N-PSDB; AA50661.
XX
XX CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
XX SIV.
XX
XX Example 1; Col 29-42; 39pp; English.
XX
XX The present sequence is that of fusion protein CD4Egammal comprising the
XX extracellular portion of CD4, which binds to HIV gp120, linked at its C-
XX terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA
XX encoding CD4 was linked to IgG1 DNA at the Eap site upstream of the hinge
XX region (see AA50661). Fusion protein CD4Egammal and a nucleic acid
XX encoding it are claimed. Also claimed are a vector comprising the nucleic
XX acid, and a method of producing the fusion protein in secreted form using
XX a transformed host cell. The fusion protein may further comprise a
XX therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
XX can be administered to an animal (including humans) for treatment of HIV
XX or SIV infection, and can also be used in assays for HIV or SIV, imaging
XX and tissue stains. IgG1 fusion proteins such as CD4Egammal provide both
XX complement-mediated and cell-mediated immunity
XX
XX Sequence 631 AA;
XX
XX Query Match 95.7%; Score 3268.5; DB 3; Length 631;
XX Best Local Similarity 99.0%; Pred. No. 1.8e-170;
XX Matches 624; Conservative 0; Mismatches 1; Indels 5; Gaps 1;
XX
XX 1 MNRGVFRHLILVLAQALPAATQGNKVVYGGKGDVVELTCTASQKKSIOFMKNSNQIK 60
XX 1 MNRGVFRHLILVLAQALPAATQGNKVVYGGKGDVVELTCTASQKKSIOFMKNSNQIK 60
XX
XX 61 ILGNOSFILTGGSKNDRAADSRSLMDQGNPFLIKNLIKEDSDTYICEVEDQKEEVL 120
XX 61 ILGNOSFILTGGSKNDRAADSRSLMDQGNPFLIKNLIKEDSDTYICEVEDQKEEVL 120
XX
XX 61 ILGNOSFILTGGSKNDRAADSRSLMDQGNPFLIKNLIKEDSDTYICEVEDQKEEVL 120
XX 61 ILGNOSFILTGGSKNDRAADSRSLMDQGNPFLIKNLIKEDSDTYICEVEDQKEEVL 120
XX
XX 121 LVFGLTANSDTHLLOQSITLTLESPPGSSPVQCSPPGKNIQGGKTLVSQLELDQSG 180
XX 121 LVFGLTANSDTHLLOQSITLTLESPPGSSPVQCSPPGKNIQGGKTLVSQLELDQSG 180
XX
XX 121 LVFGLTANSDTHLLOQSITLTLESPPGSSPVQCSPPGKNIQGGKTLVSQLELDQSG 180
XX 121 LVFGLTANSDTHLLOQSITLTLESPPGSSPVQCSPPGKNIQGGKTLVSQLELDQSG 180
XX
XX 181 TWTCTVLNQKQKVEFKIDIVLAFOKASSIVYKKEGEQVFSFPLAFVTEKLTGSGELMW 240
XX 181 TWTCTVLNQKQKVEFKIDIVLAFOKASSIVYKKEGEQVFSFPLAFVTEKLTGSGELMW 240
XX
XX 181 TWTCTVLNQKQKVEFKIDIVLAFOKASSIVYKKEGEQVFSFPLAFVTEKLTGSGELMW 240
XX 181 TWTCTVLNQKQKVEFKIDIVLAFOKASSIVYKKEGEQVFSFPLAFVTEKLTGSGELMW 240
XX
XX 241 QAEBAASSSKWITFDLKNKEVSVKRTOPKIQMGKKLHLHTLPPALQYVAGSGNLTIA 300
XX 241 QAEBAASSSKWITFDLKNKEVSVKRTOPKIQMGKKLHLHTLPPALQYVAGSGNLTIA 300
XX
XX 241 QAEBAASSSKWITFDLKNKEVSVKRTOPKIQMGKKLHLHTLPPALQYVAGSGNLTIA 300
XX 241 QAEBAASSSKWITFDLKNKEVSVKRTOPKIQMGKKLHLHTLPPALQYVAGSGNLTIA 300
XX
XX 301 LEAKTGKLHQEVNLVVMRATQLOKNLTCEVMGPTSPKMLSLKLEKKAIVSKREKPVWV 360
XX

```

```

DB 301 LEAKTGKLHQEVNLVVMRATQLOKNLTCEVMGPTSPKMLSLKLEKKAIVSKREKPVWV 360
QY 361 LNPENGMQCLISDSGOVLBSNIVLPTWSTPV-----EPKSCKTHTCPPCPAPELLG 415
DB 361 LNPENGMQCLISDSGOVLBSNIVLPTWSTPVHADDEEPKSCDKTHTCPCPAPELLG 420
QY 416 GPSVFLPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDQVFNHNAKTKREBOY 475
DB 421 GPSVFLPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKFNMYVDQVFNHNAKTKREBOY 480
QY 476 NSTYRVSVLTIVLHODMNGKEKCKVSNKALPAPIEKTISKAKGQPREPPVYTLPSPRD 535
DB 481 NSTYRVSVLTIVLHODMNGKEKCKVSNKALPAPIEKTISKAKGQPREPPVYTLPSPRD 540
QY 536 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 595
DB 541 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSR 600
QY 596 WQGGNVFSCSVMHREALHNNHTYQKSLSLSPG 625
DB 601 WQGGNVFSCSVMHREALHNNHTYQKSLSLSPG 630
XX
XX RESULT 3
XX AA51079
XX ID AA51079 standard; protein; 631 AA.
XX
XX AA51079;
XX
XX 23-MAR-2000 (first entry)
XX
XX Human fusion protein CD4Egammal.
XX
XX Fusion protein; human; CD4; IgG1; immunoglobulin; gp120;
XX anti-human immunodeficiency virus; CD4Egammal.
XX
XX OS Homo sapiens.
XX Synthetic.
XX
XX US6004781-A.
XX
XX 21-DEC-1999.
XX
XX 04-FEB-1994; 94US-00191708.
XX
XX 22-JAN-1988; 88US-00147351.
XX 23-JAN-1989; 89US-00299596.
XX 09-JUN-1992; 92US-00896781.
XX 12-APR-1993; 93US-00057952.
XX
XX (GENO ) GEN HOSPITAL CORP.
XX
XX Seed B;
XX
XX WPI; 2000-085792/07.
XX N-PSDB; AA244062.
XX
XX Fusion protein useful for the treatment of human immunodeficiency virus.
XX
XX Example 1; Col 29-42; 39pp; English.
XX
XX This invention describes a novel nucleic acid (I) encoding a fusion
XX protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
XX and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
XX chain (III). The products of the invention have anti-human
XX immunodeficiency virus (HIV) activity and are capable of binding to
XX gp120. The fusion protein is useful for treating human immunodeficiency
XX virus (HIV) or simian immunodeficiency virus (SIV). This sequence
XX represents the fusion protein CD4Egammal which is constructed from CD4
XX linked to human IgG1 upstream of the hinge region
XX
XX Sequence 631 AA;
XX

```



Query Match 95.7%; Score 3268.5; DB 3; Length 631;  
Best Local Similarity 99.0%; Pred. No. 1.8e-170;  
Matches 624; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

```
QY 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 60
DB 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 60
QY 61 ILNGQSFLLTKGPSKLNDRADSRSLMDQGNPFLIKLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGQSFLLTKGPSKLNDRADSRSLMDQGNPFLIKLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLQNGKKEFKIDIVLAFQKASSIYKKEGQVFSPFLATFVEKLTGSGELMW 240
DB 181 TWTCVTLQNGKKEFKIDIVLAFQKASSIYKKEGQVFSPFLATFVEKLTGSGELMW 240
QY 241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPOALPOYAGSGNLTLLA 300
DB 241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPOALPOYAGSGNLTLLA 300
QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNTLCEVWGPTSPKLMLSLKLENKEAVSKREKPYW 360
DB 301 LEAKTGKLGHOEVNLVVMRATOLQKNTLCEVWGPTSPKLMLSLKLENKEAVSKREKPYW 360
QY 361 LNPEAGMMOCLLSDSGVLLSNIKYVLTPTWSTPV-----EPKSCDKHTHTCPCPAPELLG 415
DB 361 LNPEAGMMOCLLSDSGVLLSNIKYVLTPTWSTPVADDEBPSCDKHTHTCPCPAPELLG 420
QY 416 GPSVFLPPPKPDITMISRTPEVTCVVVDVSHEDPEVKFNNYVVGVEVHNAKTKPREEQY 475
DB 421 GPSVFLPPPKPDITMISRTPEVTCVVVDVSHEDPEVKFNNYVVGVEVHNAKTKPREEQY 480
QY 476 NSTYRVSVLTITVLDMLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRD 535
DB 481 NSTYRVSVLTITVLDMLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRD 540
QY 536 ELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSR 595
DB 541 ELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSR 600
QY 596 WOQGNVFSCSVMHEALHNHYTQKSLSLSPG 625
DB 601 WOQGNVFSCSVMHEALHNHYTQKSLSLSPG 630
```

RESULT 4  
AA59169  
AA59169 standard; protein; 631 AA.

AA59169;  
14-MAR-2000 (first entry)

CD4-Ig fusion protein CD4Bgamma1.

HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;  
secreted protein; HIV infection; medicament.

Synthetic.

Homo sapiens.

CA1340741-C.

14-SEP-1999.

20-JAN-1989; 89CA-00588749.

20-JAN-1989; 89CA-00588749.

XX (GENO) GEN HOSPITAL CORP.

XX Seed B;

XX WPI; 2000-063015/06.

XX N-PSDB; AA248202.

PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in  
the treatment of HIV or simian immunodeficiency virus infections.

PS Example 1; Page 37-46; 89pp; English.

CC The invention provides a fusion gene encoding a fusion protein that  
comprises an extracellular CD4 DNA sequence or its fragment which binds  
to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA  
sequence of an Ig heavy or light chain, where the DNA sequence encoding  
the variable region has been replaced with the DNA sequence which encodes  
extracellular CD4 or its gp120 binding fragment. The fusion protein is  
capable of being secreted. The fusion proteins are useful for treating  
HIV or SIV infections in animals, preferably humans. They are also useful  
for producing medicaments which can be used for creating HIV or SIV  
infections in humans. The present sequence represents the fusion protein  
CD4Bgamma1 where the CD4 is linked to human IgG1 at the Esp site upstream  
of the hinge region

XX Sequence 631 AA:

Query Match 95.7%; Score 3268.5; DB 3; Length 631;  
Best Local Similarity 99.0%; Pred. No. 1.8e-170;  
Matches 624; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

```
QY 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 60
DB 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 60
QY 61 ILNGQSFLLTKGPSKLNDRADSRSLMDQGNPFLIKLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGQSFLLTKGPSKLNDRADSRSLMDQGNPFLIKLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLTLSPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLQNGKKEFKIDIVLAFQKASSIYKKEGQVFSPFLATFVEKLTGSGELMW 240
DB 181 TWTCVTLQNGKKEFKIDIVLAFQKASSIYKKEGQVFSPFLATFVEKLTGSGELMW 240
QY 241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPOALPOYAGSGNLTLLA 300
DB 241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPOALPOYAGSGNLTLLA 300
QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNTLCEVWGPTSPKLMLSLKLENKEAVSKREKPYW 360
DB 301 LEAKTGKLGHOEVNLVVMRATOLQKNTLCEVWGPTSPKLMLSLKLENKEAVSKREKPYW 360
QY 361 LNPEAGMMOCLLSDSGVLLSNIKYVLTPTWSTPV-----EPKSCDKHTHTCPCPAPELLG 415
DB 361 LNPEAGMMOCLLSDSGVLLSNIKYVLTPTWSTPVADDEBPSCDKHTHTCPCPAPELLG 420
QY 416 GPSVFLPPPKPDITMISRTPEVTCVVVDVSHEDPEVKFNNYVVGVEVHNAKTKPREEQY 475
DB 421 GPSVFLPPPKPDITMISRTPEVTCVVVDVSHEDPEVKFNNYVVGVEVHNAKTKPREEQY 480
QY 476 NSTYRVSVLTITVLDMLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRD 535
DB 481 NSTYRVSVLTITVLDMLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRD 540
QY 536 ELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSR 595
DB 541 ELTKNQVSLTCLVKGFYPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSR 600
QY 596 WOQGNVFSCSVMHEALHNHYTQKSLSLSPG 625
```



DB	601	WQGNVVFSCSVHNEALHNHYTOKSLSLISFG	630
RESULT 5			
ID	AAP93008		
XX	AAP93008	standard; protein; 729 AA.	
XX			
AC	AAP93008;		
XX			
DT	25-MAR-2003	(revised)	
DT	02-NOV-1992	(first entry)	
XX			
DE	Genetic construct which encodes CD4 linked to human IgG1 at the Hind3		
DE	site upstream of the CH1 region (fusion protein CD4H-gamma-1).		
XX			
KV	Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;		
XX	diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.		
OS	Homo sapiens.		
PN	EP325262-A.		
PD	26-JUL-1989.		
XX			
PF	20-JAN-1989; 89EP-00100913.		
XX			
PR	22-JAN-1988; 88US-00147351.		
PA	(GEHO ) GEN HOSPITAL CORP.		
XX			
PI	Seed B;		
DR	WP1; 1989-214472/30.		
DR	N-PSDB; AAN90356.		
XX			
PT	Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV		
XX	infections or detecting HIV or SIV in sample.		
XX			
XX	Example; Table 1, Page 12-23; 68pp; English.		
CC	The fusion protein genes of the invention pref. comprise cDNA sequences		
CC	which encode CD4 or a fragment which binds gp120 ligated to an expression		
CC	plasmid which encodes an antibody in which the variable region of the		
CC	gene has been deleted (see WO87-02671). The CD4 portion of the fusion		
CC	protein may comprise the complete CD4 sequence, the 370 AA extracellular		
CC	region and the membrane spanning domain, or the extracellular region. The		
CC	Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are		
CC	specifically claimed: fusion proteins CD4H-gamma-1, CD4mu, CD4Fmu, CD4E-		
CC	gamma1, and CD4mu (No. 67608), pCD4F-gamma (No. 67609) and pCD4E-gamma-1		
CC	(No. 67610). The plasmid containing (pCD4H-gamma-1) has been deposited in		
CC	E. coli (MC1061/B3) at the ATCC under accession number 67611. (Updated on		
CC	25-MAR-2003 to correct PA field.)		
XX			
XX	Sequence 729 AA;		
Q0			
Query Match	94.3%;	Score 3219.5;	DB 1; Length 729;
Best Local Similarity	85.7%;	Pred. No. 1e-167;	
Matches 624;	Conservative 0;	Mismatches 1;	Indels 103; Gaps 1
QY	1	MNRGVPFPHLLLVQLALLPAATGCKRVVLGKKGDTVELTCTASQKKSIOFHWKNSNQIK	60
DB	1	MNRGVPFPHLLLVQLALLPAATGCKRVVLGKKGDTVELTCTASQKKSIOFHWKNSNQIK	60
QY	61	ILNGGSPFTKPSKLTNDPADRSRLMDGQNPFLIIKNIKIEDSDYICEVEDEQKEEVL	120
DB	61	ILNGGSPFTKPSKLTNDPADRSRLMDGQNPFLIIKNIKIEDSDYICEVEDEQKEEVL	120
QY	121	LVFGTJANSPTHLLOGQSILTLLESPGSSPVSQCRSPKGNIIQGGKTLVSQLELQDSG	180
DB	121	LVFGTJANSPTHLLOGQSILTLLESPGSSPVSQCRSPKGNIIQGGKTLVSQLELQDSG	180
QY	181	TWTCITVNLQKKVERKIDIVLAFQKASSIVYKKGEOVESFPPLAFVTEKLTGSGELMW	240

Dd		181	TWCTVLQDQKKKEFIIDVLVLAFOKASSTYKKGEQEVSEFFLATVTEKLTGSGLMW	240
Qy		241	QAEASSSSKSWITFDLKNKEVSVKRVTDQPKLQMGKKLPLHLTLPOLPYAGSGNLTLA	300
Dd		241	QAEASSSSKSWITFDLKNKEVSVKRVTDQPKLQMGKKLPLHLTLPOLPYAGSGNLTLA	300
Qy		301	LEAKTQKLHGVNLVVMRATLOQKLTGCVGPGTSPKLMLSLKLENKAKVSKREKVVW	360
Dd		301	LEAKTQKLHGVNLVVMRATLOQKLTGCVGPGTSPKLMLSLKLENKAKVSKREKVVW	360
Qy		361	LNPEAGMOCCLSDSGQVLLESNIKVLPWSTP-----	393
Dd		361	LNPEAGMOCCLSDSGQVLLESNIKVLPWSTP-----	393
Qy		394	-----	393
Dd		421	TAALGCLVMDYPEPVTVSMNSGALTSGVHPPRAVLQSSGLYSLSVVTVPSSSLGTQTY	480
Qy		394	-----VEPKSCDKHTHTCPPCAPPELLCGPSVFLPPKPKDTLMTISRTPE	437
Dd		481	ICNVNHPKSTNTKYDKKVBEPKSCDKHTHTCPPCAPPELLCGPSVFLPPKPKDTLMTISRTPE	540
Qy		438	VTCVVVDVSHEDPEVKFNNYVDSGVVHNAKTRPREQVNSITRYRVSVLTLYLHODMLNGKE	497
Dd		541	VTCVVVDVSHEDPEVKFNNYVDSGVVHNAKTRPREQVNSITRYRVSVLTLYLHODMLNGKE	600
Qy		498	YKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIA	557
Dd		601	YKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNOVSLTCLVKGFPYPSDIA	660
Qy		558	VEMESNGQPENNYKTTTPVPLVSDSGSFPLYSKLTVDKSKHWQGNVFGSCVMHEALHNHYQ	617
Dd		661	VEMESNGQPENNYKTTTPVPLVSDSGSFPLYSKLTVDKSKHWQGNVFGSCVMHEALHNHYQ	720
Qy		618	KSLSLSRG 625	
Dd		721	KSLSLSPG 728	
RESULT 6				
AAB19507				
ID	AAB19507	standard; protein; 729 AA.		
XX	AC			
XX	AAB19507;			
XX	DT	09-JUN-2001 (first entry)		
DE	CD4-IgG1 fusion protein CH4Hgmam1.			
XX	CD4; IgG1; human; CD4Hgmam1; fusion protein; immunoglobulin; HIV; SIV;			
KW	gpi20; therapy; diagnosis.			
OS	Homo sapiens.			
XX	Key	Location/Qualifiers		
FT	Protein	1..395		
FT	Protein	/note="CD4 extracellular region"		
FT	Protein	400..729		
FT	Protein	/note="IgG1 heavy chain"		
XX	US6117656-A.			
PN	12-SEP-2000.			
PD	07-JUN-1995;	95US-00479353.		
PE	22-JAN-1988;	88US-00147351.		
PR	23-JAN-1989;	89US-0029596.		
PR	09-JUN-1992;	92US-00896781.		
PR	12-APR-1993;	93US-00057952.		
PR	04-FEB-1994;	94US-00191708.		
XX				



